

D4

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :  A61K 47/48, 39/00		A2	(11) International Publication Number: WO 98/36779  (43) International Publication Date: 27 August 1998 (27.08.98)
(21) International Application Number: PCT/US98/02945		(74) Agents: MAGRI, Karen, A. et al.; Myers, Bigel, Sibley & Sajovec, L.L.P., P.O. Box 37428, Raleigh, NC 27627 (US).	
(22) International Filing Date: 18 February 1998 (18.02.98)		(81) Designated States: AL, AM, AT, AT (Utility model); AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, EE (Utility model), ES, FI, FI (Utility model), GB, GE, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(30) Priority Data: 08/801,263 19 February 1997 (19.02.97) US		(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 08/801,263 (CON) Filed on 19 February 1997 (19.02.97)	
(71) Applicant (for all designated States except US): UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL [US/US]; 308 Bynum Hall, Campus Box 4105, Chapel Hill, NC 27599-4105 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): JOHNSTON, Robert, E. [US/US]; 101 Marin Place, Chapel Hill, NC 27516 (US). DAVIS, Nancy, L. [US/US]; 132 New Castle Drive, Chapel Hill, NC 27514 (US). SIMPSON, Dennis, A. [US/US]; 19A Deer Mountain Road, Pittsboro, NC 27312 (US).		Published <i>Without international search report and to be republished upon receipt of that report.</i>	

(54) Title: SYSTEM FOR THE *IN VIVO* DELIVERY AND EXPRESSION OF HETEROLOGOUS GENES IN THE BONE MARROW

## (57) Abstract

The present invention provides a method of delivering immunogenic or therapeutic proteins to bone marrow cells using alphavirus vectors. The alphavirus vectors disclosed herein target specifically to bone marrow tissue, and viral genomes persist in bone marrow for at least three months post-infection. No or very low levels of virus were detected in quadricep, brain, and sera of treated animals. The sequence of a consensus Sindbis cDNA clone, pTR339, and infectious RNA transcripts, infectious virus particles, and pharmaceutical formulations derived therefrom are also disclosed. The sequence of the genomic RNA of the Girdwood S.A. virus, and cDNA clones, infectious RNA transcripts, infectious virus particles, and pharmaceutical formulations derived therefrom are also disclosed.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Israel	UA	Ukraine
BR	Brazil	IL	Israel	MW	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	NL	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Republic of Korea	PT	Portugal		
CU	Cuba	LC	Kazakhstan	RO	Romania		
CZ	Czech Republic	LI	Saint Lucia	RU	Russian Federation		
DE	Germany	LK	Liechtenstein	SD	Sudan		
DK	Denmark	LR	Sri Lanka	SE	Sweden		
EE	Estonia		Liberia	SG	Singapore		

-1-

# SYSTEM FOR THE *IN VIVO* DELIVERY AND EXPRESSION OF HETEROLOGOUS GENES IN THE BONE MARROW

5

## FEDERALLY SPONSORED RESEARCH

This invention was made with Government support under Grant Number 5 RO1 AI22186 from the National Institutes of Health. The Government has certain rights to this invention.

10

The present invention relates to recombinant DNA technology, and in particular to introducing and expressing foreign DNA in a eukaryotic cell.

15

## BACKGROUND OF THE INVENTION

20

The Alphavirus genus includes a variety of viruses all of which are members of the Togaviridae family. The alphaviruses include Eastern Equine Encephalitis virus (EEE), Venezuelan Equine Encephalitis virus (VEE), Everglades virus, Mucambo virus, Pixuna virus, Western Equine Encephalitis virus (WEE), Sindbis virus, South African Arbovirus No. 86 (S.A.AR 86), Girdwood S.A. virus, Ockelbo virus, Semliki Forest virus, Middelburg virus, Chikungunya virus, O'Nyong-Nyong virus, Ross River virus, Barmah Forest virus, Getah virus, Sagiama virus, Bebaru virus, Mayaro virus, Una virus, Aura virus, Whataroa virus, Babanki virus, Kyzylagach virus, Highlands J virus, Fort Morgan virus, Ndumu virus, and Buggy Creek virus.

-2-

The alphavirus genome is a single-stranded, messenger-sense RNA, modified at the 5'-end with a methylated cap, and at the 3'-end with a variable-length poly (A) tract. The viral genome is divided into two regions: the first encodes the nonstructural or replicase proteins (nsP1-nsP4) and the second encodes the viral structural proteins. Strauss and Strauss, *Microbiological Rev.* 58, 491-562, 494 (1994). Structural subunits consisting of a single viral protein, C, associate with themselves and with the RNA genome in an icosahedral nucleocapsid. In the virion, the capsid is surrounded by a lipid envelope covered with a regular array of transmembranal protein spikes, each of which consists of a heterodimeric complex of two glycoproteins, E1 and E2. See Paredes et al., *Proc. Natl. Acad. Sci. USA* 90, 9095-99 (1993); Paredes et al., *Virology* 187, 324-32 (1993); Pedersen et al., *J. Virol.* 14:40 (1974).

Sindbis virus, the prototype member of the alphavirus genus of the family *Togaviridae*, and viruses related to Sindbis are broadly distributed throughout Africa, Europe, Asia, the Indian subcontinent, and Australia, based on serological surveys of humans, domestic animals and wild birds. Kokernot et al., *Trans. R. Soc. Trop. Med. Hyg.* 59, 553-62 (1965); Redaksie, *S. Afr. Med. J.* 42, 197 (1968); Adekolu-John and Fagbami, *Trans. R. Soc. Trop. Med. Hyg.* 77, 149-51 (1983); Darwish et al., *Trans. R. Soc. Trop. Med. Hyg.* 77, 442-45 (1983); Lundström et al., *Epidemiol. Infect.* 106, 567-74 (1991); Morrill et al., *J. Trop. Med. Hyg.* 94, 166-68 (1991). The first isolate of Sindbis virus (strain AR339) was recovered from a pool of *Culex* sp. mosquitoes collected in Sindbis, Egypt in 1953 (Taylor et al., *Am. J. Trop. Med. Hyg.* 4, 844-62 (1955)), and is the most extensively studied representative of this group. Other members of the Sindbis group of alphaviruses include South African Arbovirus No. 86, Ockelbo82, and Girdwood S.A. These viruses are not strains of the Sindbis virus; they are related to Sindbis AR339, but they are more closely related to each other based on nucleotide sequence and serological comparisons. Lundström et al., *J. Wildl. Dis.* 29, 189-95 (1993); Simpson et al., *Virology* 222, 464-69 (1996). Ockelbo82, S.A.AR86 and Girdwood S.A. are all associated with human disease, whereas Sindbis is not. The clinical symptoms of human infection with Ockelbo82,

-3-

S.A.AR86, or Girdwood S.A. are a febrile illness, general malaise, macropapular rash, and joint pain that occasionally progresses to a polyarthralgia sometimes lasting from a few months to a few years.

5           The study of these viruses has led to the development of beneficial techniques for vaccinating against the alphavirus diseases, and other diseases through the use of alphavirus vectors for the introduction of foreign DNA. *See* United States Patent No. 5,185,440 to Davis et al., and PCT Publication WO 92/10578. It is intended that all United States patent references be incorporated in their entirety by reference.

10          It is well known that live, attenuated viral vaccines are among the most successful means of controlling viral disease. However, for some virus pathogens, immunization with a live virus strain may be either impractical or unsafe. One alternative strategy is the insertion of sequences encoding immunizing antigens of such agents into a vaccine strain of another virus. One such system  
15 utilizing a live VEE vector is described in United States Patent No. 5,505,947 to Johnston et al.

20          Sindbis virus vaccines have been employed as viral carriers in virus constructs which express genes encoding immunizing antigens for other viruses. *See* United States Patent No. 5,217,879 to Huang et al. Huang et al. describes Sindbis infectious viral vectors. However, the reference does not describe the cDNA sequence of Girdwood S.A. and TR339, nor clones or viral vectors produced therefrom.

25          Another such system is described by Hahn et al., *Proc. Natl. Acad. Sci. USA* 89:2679 (1992), wherein Sindbis virus constructs which express a truncated form of the influenza hemagglutinin protein are described. The constructs are used to study antigen processing and presentation *in vitro* and in mice. Although no infectious challenge dose is tested, it is also suggested that

-4-

such constructs might be used to produce protective B- and T-cell mediated immunity.

5                   London et al., *Proc. Natl. Acad. Sci. USA* 89, 207-11 (1992), disclose a method of producing an immune response in mice against a lethal Rift Valley Fever (RVF) virus by infecting the mice with an infectious Sindbis virus containing an RVF epitope. London does not disclose using Girdwood S.A. or TR339 to induce an immune response in animals.

10                  Viral carriers can also be used to introduce and express foreign DNA in eukaryotic cells. One goal of such techniques is to employ vectors that target expression to particular cells and/or tissues. A current approach has been to remove target cells from the body, culture them *ex vivo*, infect them with an expression vector, and then reintroduce them into the patient.

15                  PCT Publication No. WO 92/10578 to Garoff and Liljeström provide a system for introducing and expressing foreign proteins in animal cells using alphaviruses. This reference discloses the use of Semliki Forest virus to introduce and express foreign proteins in animal cells. The use of Girdwood S.A. or TR339 is not discussed. Furthermore, this reference does not provide a method of targeting and introducing foreign DNA into specific cell or tissue types.

20                  Accordingly, there remains a need in the art for full-length cDNA clones of positive-strand RNA viruses, such as Girdwood S.A and TR339. In addition, there is an ongoing need in the art for improved vaccination strategies. Finally, there remains a need in the art for improved methods and nucleic acid sequences for delivering foreign DNA to target cells.

#### SUMMARY OF THE INVENTION

25                  A first aspect of the present invention is a method of introducing and expressing heterologous RNA in bone marrow cells, comprising: (a) providing

-5-

5 a recombinant alphavirus, the alphavirus containing a heterologous RNA segment, the heterologous RNA segment comprising a promoter operable in bone marrow cells operatively associated with a heterologous RNA to be expressed in bone marrow cells; and then (b) contacting the recombinant alphavirus to the bone marrow cells so that the heterologous RNA segment is introduced and expressed therein.

10 As a second aspect, the present invention provides a helper cell for expressing an infectious, propagation defective, Girdwood S.A. virus particle, comprising, in a Girdwood S.A.-permissive cell: (a) a first helper RNA encoding (i) at least one Girdwood S.A. structural protein, and (ii) not encoding at least one other Girdwood S.A. structural protein; and (b) a second helper RNA separate from the first helper RNA, the second helper RNA (i) not encoding the at least one Girdwood S.A. structural protein encoded by the first helper RNA, and (ii) encoding the at least one other Girdwood S.A. structural protein not encoded by 15 the first helper RNA, and with all of the Girdwood S.A. structural proteins encoded by the first and second helper RNAs assembling together into Girdwood S.A. particles in the cell containing the replicon RNA; and wherein the Girdwood S.A. packaging segment is deleted from at least the first helper RNA.

20 A third aspect of the present invention is a method of making infectious, propagation defective, Girdwood S.A. virus particles, comprising: transfecting a Girdwood S.A.-permissive cell with a propagation defective replicon RNA, the replicon RNA including the Girdwood S.A. packaging segment and an inserted heterologous RNA; producing the Girdwood S.A. virus particles in the transfected cell; and then collecting the Girdwood S.A. virus particles from the 25 cell. Also disclosed are infectious Girdwood S.A. RNAs, cDNAs encoding the same, infectious Girdwood S.A. virus particles, and pharmaceutical formulations thereof.

As a fourth aspect, the present invention provides a helper cell for expressing an infectious, propagation defective, TR339 virus particle, comprising,

-6-

in a TR339-permissive cell: (a) a first helper RNA encoding (i) at least one TR339 structural protein, and (ii) not encoding at least one other TR339 structural protein; and (b) a second helper RNA separate from the first helper RNA, the second helper RNA (i) not encoding the at least one TR339 structural protein encoded by the first helper RNA, and (ii) encoding the at least one other TR339 structural protein not encoded by the first helper RNA, and with all of the TR339 structural proteins encoded by the first and second helper RNAs assembling together into TR339 particles in the cell containing the replicon RNA; and wherein the TR339 packaging segment is deleted from at least the first helper RNA.

A fifth aspect of the present invention is a method of making infectious, propagation defective, TR339 virus particles, comprising: transfecting a TR339-permissive cell with a propagation defective replicon RNA, the replicon RNA including the TR339 packaging segment and an inserted heterologous RNA; producing the TR339 virus particles in the transfected cell; and then collecting the TR339 virus particles from the cell. Also disclosed are infectious TR339 RNAs, cDNAs encoding the same, infectious TR339 virus particles, and pharmaceutical formulations thereof.

As a sixth aspect, the present invention provides a recombinant DNA comprising a cDNA coding for an infectious Girdwood S.A. virus RNA transcript, and a heterologous promoter positioned upstream from the cDNA and operatively associated therewith. The present invention also provides infectious RNA transcripts encoded by the above-mentioned cDNA and infectious viral particles containing the infectious RNA transcripts.

As a seventh aspect, the present invention provides a recombinant DNA comprising a cDNA coding for a Sindbis strain TR339 RNA transcript, and a heterologous promoter positioned upstream from the cDNA and operatively associated therewith. The present invention also provides infectious RNA transcripts encoded by the above-mentioned cDNA and infectious viral particles containing the infectious RNA transcripts.

The foregoing and other aspects of the present invention are described in the detailed description set forth below.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 presents the cDNA sequence (SEQ ID NO:1) of S.A.AR86. The RNA sequence of the 5' 40 nucleotides was obtained by direct sequencing of the genomic RNA. The rest of the genome was sequenced by RT-PCR of fragments amplified from virion RNA. Nucleotides 1 through 59 represent the 5' UTR, the non-structural polyprotein is encoded by nucleotides 60 through 7559 (nsP1--nt60 through nt1679; nsP2--nt1680 through nt4099; nsP3--nt4100 through nt5729; nsP4--nt5730 through nt7559), the structural polyprotein is encoded by nucleotides 7608 through 11342 (capsid--nt7608 through nt8399; E3--nt8400 through nt8591; E2--nt8592 through nt9860; 6K--nt9861 through nt10025; E1--nt10026 through nt11342), and the 3' UTR is represented by nucleotides 11346 through 11663.

Figure 1A shows nucleotides 1 through 3800 of the cDNA sequence of S.A.AR86.

Figure 1B shows nucleotides 3801 through 7900 of the cDNA sequence of S.A.AR86.

Figure 1C shows nucleotides 7901 through 11663 of the cDNA sequence of S.A.AR86.

Figure 2 presents the putative amino acid sequences of the S.A.AR86 polyproteins (SEQ ID NO:2 and SEQ ID NO:3). The amino acids were derived from the S.A.AR86 cDNA sequence given in Figure 1 (SEQ ID NO:1).

-8-

Figure 2A shows the amino acid sequence of the non-structural polyprotein of S.A.AR86 (SEQ ID NO:2).

Figure 2B shows the amino acid sequence of the structural polyprotein of S.A.AR86 (SEQ ID NO:3).

5       Figure 3 presents the cDNA sequence (SEQ ID NO:4) of Girdwood S.A. The RNA sequence of the 5' 40 nucleotides was obtained by direct sequencing of the genomic RNA. The rest of the genome sequence was obtained by sequencing of fragments amplified by RT-PCR from virion RNA. An "N" in the sequence indicates that the identity of the nucleotide at that position is unknown. Nucleotides 1 through 59 represent the 5' UTR, the non-structural polyprotein is encoded by nucleotides 60 through 7613 (nsP1-nt60 through nt1679; nsP2-nt1680 through nt4099; nsP3-nt4100 through nt5762 or nt5783; nsP4-nt5784 through nt7613), the structural polyprotein is encoded by nucleotides 7662 through 11396 (capsid-nt7662 through nt8453; E3-nt8454 through nt8645; E2-nt8646 through nt9914, 6K-9915 through nt10079; E1-nt10080 through nt11396), and the 3' UTR is represented by nucleotides 11400 through 11717. There is an opal termination codon at nucleotides 5763 through 5765.

10

15

Figure 3A shows nucleotides 1 through 3800 of the cDNA sequence of Girdwood S.A.

20       Figure 3B shows nucleotides 3801 through 7900 of the cDNA sequence of Girdwood S.A.

Figure 3C shows nucleotides 7901 through 11717 of the cDNA sequence of Girdwood S.A.

25       Figure 4 illustrates the putative amino acid sequences of the Girdwood S.A. polyproteins (SEQ ID NO:5 and SEQ ID NO:6). The amino

-9-

acids were derived from the Girdwood S.A. cDNA sequence given in Figure 3 (SEQ ID NO:4).

5           Figure 4A shows the amino acid sequence of the non-structural polyprotein of Girdwood S.A. The sequence terminates at the opal termination codon. The complete amino acid sequence is presented in SEQ ID NO:5.

Figure 4B shows the amino acid sequence of the structural polyprotein of Girdwood S.A. (SEQ ID NO:6).

Figure 5 illustrates the nucleotide sequence (SEQ ID NO:7) of clone pS55, a cDNA clone of the S.A.AR86 genomic RNA.

10           Figure 5A shows nucleotides 1 through 6720 of the cDNA sequence of pS55.

Figure 5B shows nucleotides 6721 through 11663 of the cDNA sequence of pS55.

15           Figure 6 presents the cDNA sequence (SEQ ID NO:8) of clone pTR339. The TR339 virus is derived from this clone. Nucleotides 1 through 59 represent the 5' UTR, the non-structural polyprotein is encoded by nucleotides 60 through 7598 (nsP1--nt60 through nt1679; nsP2--nt1680 through nt4099; nsP3--nt4100 through nt5747 or 5768; nsP4--nt5769 through nt7598), the structural polyprotein is encoded by nucleotides 7647 through 11381 (capsid--nt7647 through nt8438; E3--nt8439 through nt8630; E2--nt8631 through nt9899; 6K--nt9900 through nt10064; E1--nt10065 through nt11381), and the 3' UTR is represented by nucleotides 11382 through 11703. There is an opal termination codon at nucleotides 5748 through 5750.

25           Figure 6A shows nucleotides 1 through 6720 of the cDNA sequence of pTR339.

-10-

Figure 6B shows nucleotides 6721 through 11703 of the cDNA sequence of pTR339.

DETAILED DESCRIPTION OF THE INVENTION

The production and use of recombinant DNA, vectors, transformed host cells, selectable markers, proteins, and protein fragments by genetic engineering are well-known to those skilled in the art. See, e.g., United States Patent No. 4,761,371 to Bell et al. at Col. 6 line 3 to Col. 9 line 65; United States Patent No. 4,877,729 to Clark et al. at Col. 4 line 38 to Col. 7 line 6; United States Patent No. 4,912,038 to Schilling at Col 3 line 26 to Col 14 line 12; and United States Patent No. 4,879,224 to Wallner at Col. 6 line 8 to Col. 8 line 59.

The term "alphavirus" has its conventional meaning in the art, and includes the various species of alphaviruses such as Eastern Equine Encephalitis virus (EEE), Venezuelan Equine Encephalitis virus (VEE), Everglades virus, Mucambo virus, Pixuna virus, Western Encephalitis virus (WEE), Sindbis virus, South African Arbovirus No. 86, Girdwood S.A. virus, Ockelbo virus, Semliki Forest virus, Middelburg virus, Chikungunya virus, O'Nyong-Nyong virus, Ross River virus, Barmah Forest virus, Getah virus, Sagiyama virus, Bebaru virus, Mayaro virus, Una virus, Aura virus, Whataroa virus, Babanki virus, Kyzlagach virus, Highlands J virus, Fort Morgan virus, Ndumu virus, Buggy Creek virus, and any other virus classified by the International Committee on Taxonomy of Viruses (ICTV) as an alphavirus. The preferred alphaviruses for use in the present invention include Sindbis virus strains (e.g., TR339), Girdwood S.A., S.A.AR86, and Ockelbo82.

An "Old World alphavirus" is a virus that is primarily distributed throughout the Old World. Alternately stated, an Old World alphavirus is a virus that is primarily distributed throughout Africa, Asia, Australia and New Zealand, or Europe. Exemplary Old World viruses include SF group alphaviruses and SIN group alphaviruses. SF group alphaviruses include Semliki Forest virus, Middelburg virus, Chikungunya virus, O'Nyong-Nyong virus, Ross River virus,

-11-

Barmah Forest virus, Getah virus, Sagiyma virus, Bebaru virus, Mayaro virus, and Una virus. SIN group alphaviruses include Sindbis virus, South African Arbovirus No. 86, Ockelbo virus, Girdwood S.A. virus, Aura virus, Whataroa virus, Babanki virus, and Kyzylagach virus.

5           Acceptable alphaviruses include those containing attenuating mutations. The phrases "attenuating mutation" and "attenuating amino acid," as used herein, mean a nucleotide sequence containing a mutation, or an amino acid encoded by a nucleotide sequence containing a mutation, which mutation results in a decreased probability of causing disease in its host (*i.e.*, a loss of virulence),  
10           in accordance with standard terminology in the art, whether the mutation be a substitution mutation or an in-frame deletion mutation. *See, e.g.*, B. DAVIS ET AL., MICROBIOLOGY 132 (3d ed. 1980). The phrase "attenuating mutation" excludes mutations or combinations of mutations which would be lethal to the virus.

15           Appropriate attenuating mutations will be dependent upon the alphavirus used. Suitable attenuating mutations within the alphavirus genome will be known to those skilled in the art. Exemplary attenuating mutations include, but are not limited to, those described in United States Patent No. 5,505,947 to Johnston et al., copending United States application 08/448,630 to Johnston et al.,  
20           and copending United States application 08/446,932 to Johnston et al. It is intended that all United States patent references be incorporated in their entirety by reference.

25           Attenuating mutations may be introduced into the RNA by performing site-directed mutagenesis on the cDNA which encodes the RNA, in accordance with known procedures. *See, Kunkel, Proc. Natl. Acad. Sci. USA* 82, 488 (1985), the disclosure of which is incorporated herein by reference in its entirety. Alternatively, mutations may be introduced into the RNA by replacement of homologous restriction fragments in the cDNA which encodes for the RNA, in accordance with known procedures.

I. Methods for Introducing and Expressing Heterologous RNA in Bone Marrow Cells.

The present invention provides methods of using a recombinant alphavirus to introduce and express a heterologous RNA in bone marrow cells. Such methods are useful as vaccination strategies when the heterologous RNA encodes an immunogenic protein or peptide. Alternatively, such methods are useful in introducing and expressing in bone marrow cells an RNA which encodes a desirable protein or peptide, for example, a therapeutic protein or peptide.

The present invention is carried out using a recombinant alphavirus to introduce a heterologous RNA into bone marrow cells. Any alphavirus that targets and infects bone marrow cells is suitable. Preferred alphaviruses include Old World alphaviruses, more preferably SF group alphaviruses and SIN group alphaviruses, more preferably Sindbis virus strains (e.g., TR339), S.A.AR86 virus, Girdwood S.A. virus, and Ockelbo virus. In a more preferred embodiment, the alphavirus contains one or more attenuating mutations, as described hereinabove.

Two types of recombinant virus vector are contemplated in carrying out the present invention. In one embodiment employing "double promoter vectors," the heterologous RNA is inserted into a replication and propagation competent virus. Double promoter vectors are described in United States Patent No. 5,505,947 to Johnston et al. With this type of viral vector, it is preferable that heterologous RNA sequences of less than 3 kilobases are inserted into the viral vector, more preferably those less than 2 kilobases, and more preferably still those less than 1 kilobase. In an alternate embodiment, propagation-defective "replicon vectors," as described in copending United States application 08/448,630 to Johnston et al., will be used. One advantage of replicon viral vectors is that larger RNA inserts, up to approximately 4-5 kilobases in length can be utilized. Double promoter vectors and replicon vectors are described in more detail hereinbelow.

-13-

The recombinant alphaviruses of the claimed method target the heterologous RNA to bone marrow cells, where it expresses the encoded protein or peptide. Heterologous RNA can be introduced and expressed in any cell type found in the bone marrow. Bone marrow cells that may be targeted by the recombinant alphaviruses of 5 the present invention include, but are not limited to, polymorphonuclear cells, hemopoietic stem cells (including megakaryocyte colony forming units (CFU-M), spleen colony forming units (CFU-S), erythroid colony forming units (CFU-E), erythroid burst forming units (BFU-E), and colony forming units in culture (CFU-C), erythrocytes, macrophages (including reticular cells), monocytes, granulocytes, megakaryocytes, lymphocytes, 10 fibroblasts, osteoprogenitor cells, osteoblasts, osteoclasts, marrow stromal cells, chondrocytes and other cells of synovial joints. Preferably, marrow cells within the endosteum are targeted, more preferably osteoblasts. Also preferred are methods in which cells in the endosteum of synovial joints (e.g., hip and knee joints) are targeted.

By targeting to the cells of the bone marrow, it is meant that the primary 15 site in which the virus will be localized *in vivo* is the cells of the bone marrow. Alternately stated, the alphaviruses of the present invention target bone marrow cells, such that titers in bone marrow two days after infection are greater than 100 PFU/g crushed bone, preferably greater than 200 PFU/g crushed bone, more preferably greater than 300 PFU/g crushed bone, and more preferably still greater than 500 PFU/g crushed bone. 20 Virus may be detected occasionally in other cell or tissue types, but only sporadically and usually at low levels. Virus localization in the bone marrow can be demonstrated by any suitable technique known in the art, such as *in situ* hybridization.

Bone marrow cells are long-lived and harbor infectious alphaviruses for a prolonged period of time, as demonstrated in the Examples below. These characteristics 25 of bone marrow cells render the present invention useful not only for the purpose of supplying a desired protein or peptide to skeletal tissue, but also for expressing proteins or peptides *in vivo* that are needed by other cell or tissue types.

The present invention can be carried out *in vivo* or with cultured bone marrow cells *in vitro*. Bone marrow cell cultures include primary cultures

-14-

of bone marrow cells, serially-passaged cultures of bone marrow cells, and cultures of immortalized bone marrow cell lines. Bone marrow cells may be cultured by any suitable means known in the art.

5       The recombinant alphaviruses of the present invention carry a heterologous RNA segment. The heterologous RNA segment encodes a promoter and an inserted heterologous RNA. The inserted heterologous RNA may encode any protein or a peptide which is desirably expressed by the host bone marrow cells. Suitable heterologous RNA may be of prokaryotic (e.g., RNA encoding the *Botulinus* toxin C), or eukaryotic (e.g., RNA encoding malaria *Plasmodium* protein cs1) origin. Illustrative proteins and peptides encoded by the heterologous RNAs of the present invention include hormones, growth factors, interleukins, cytokines, chemokines, enzymes, and ribozymes. Alternately, the heterologous RNAs encode any therapeutic protein or peptide. As a further alternative, the heterologous RNAs of the present invention encode any immunogenic protein or peptide.

10

15

An immunogenic protein or peptide, or "immunogen," may be any protein or peptide suitable for protecting the subject against a disease, including but not limited to microbial, bacterial, protozoal, parasitic, and viral diseases. For example, the immunogen may be an orthomyxovirus immunogen (e.g., an influenza virus immunogen, such as the influenza virus hemagglutinin (HA) surface protein or the influenza virus nucleoprotein gene, or an equine influenza virus immunogen), or a lentivirus immunogen (e.g., an equine infectious anemia virus immunogen, a Simian Immunodeficiency Virus (SIV) immunogen, or a Human Immunodeficiency Virus (HIV) immunogen, such as the HIV envelope 20 GP160 protein and the HIV matrix/capsid proteins). The immunogen may also be an arenavirus immunogen (e.g., Lassa fever virus immunogen, such as the Lassa fever virus nucleocapsid protein gene and the Lassa fever envelope glycoprotein gene), a poxvirus immunogen (e.g., vaccinia), a flavivirus immunogen (e.g., a yellow fever virus immunogen or a Japanese encephalitis virus immunogen), a 25 filovirus immunogen (e.g., an Ebola virus immunogen, or a Marburg virus 30

-15-

5 immunogen), a bunyavirus immunogen (*e.g.*, RVFV, CCHF, and SFS viruses), or a coronavirus immunogen (*e.g.*, an infectious human coronavirus immunogen, such as the human coronavirus envelope glycoprotein gene, or a transmissible gastroenteritis virus immunogen for pigs, or an infectious bronchitis virus immunogen for chickens).

10 Alternatively, the present invention can be used to express heterologous RNAs encoding antisense oligonucleotides. In general, "antisense" refers to the use of small, synthetic oligonucleotides to inhibit gene expression by inhibiting the function of the target mRNA containing the complementary sequence. Milligan, J.F. et al., *J. Med. Chem.* 36(14), 1923-1937 (1993). Gene expression is inhibited through hybridization to coding (sense) sequences in a specific mRNA target by hydrogen bonding according to Watson-Crick base pairing rules. The mechanism of antisense inhibition is that the exogenously applied oligonucleotides decrease the mRNA and protein levels of the target gene.

15 Milligan, J.F. et al., *J. Med. Chem.* 36(14), 1923-1937 (1993). *See also* Helene, C. and Toulme, J., *Biochim. Biophys. Acta* 1049, 99-125 (1990); Cohen, J.S., Ed., OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION, CRC Press:Boca Raton, FL (1987).

20 Antisense oligonucleotides may be of any suitable length, depending on the particular target being bound. The only limits on the length of the antisense oligonucleotide is the capacity of the virus for inserted heterologous RNA. Antisense oligonucleotides may be complementary to the entire mRNA transcript of the target gene or only a portion thereof. Preferably the antisense oligonucleotide is directed to an mRNA region containing a junction between 25 intron and exon. Where the antisense oligonucleotide is directed to an intron/exon junction, it may either entirely overlie the junction or may be sufficiently close to the junction to inhibit splicing out of the intervening exon during processing of precursor mRNA to mature mRNA (*e.g.*, with the 3' or 5' terminus of the antisense oligonucleotide being positioned within about, for example, 10, 5, 3 or

-16-

2 nucleotides of the intron/exon junction). Also preferred are antisense oligonucleotides which overlap the initiation codon.

When practicing the present invention, the antisense oligonucleotides administered may be related in origin to the species to which it is administered.  
5 When treating humans, human antisense may be used if desired.

Promoters for use in carrying out the present invention are operable in bone marrow cells. An operable promoter in bone marrow cells is a promoter that is recognized by and functions in bone marrow cells. Promoters for use with the present invention must also be operatively associated with the heterologous RNA to be expressed in the bone marrow. A promoter is operably linked to a heterologous RNA if it controls the transcription of the heterologous RNA, where the heterologous RNA comprises a coding sequence. Suitable promoters are well known in the art. The Sindbis 26S promoter is preferred when the alphavirus is a strain of Sindbis virus. Additional preferred promoters beyond the Sindbis 10 26S promoter include the Girdwood S.A. 26S promoter when the alphavirus is Girdwood S.A., the S.A.AR86 26S promoter when the alphavirus is S.A.AR86, and any other promoter sequence recognized by alphavirus polymerases. Alphavirus promoter sequences containing mutations which alter the activity level 15 of the promoter (in relation to the activity level of the wild-type) are also suitable in the practice of the present invention. Such mutant promoter sequences are described in Raju and Huang, *J. Virol.* 65, 2501-2510 (1991), the disclosure of 20 which is incorporated in its entirety by reference.

The heterologous RNA is introduced into the bone marrow cells by contacting the recombinant alphavirus carrying the heterologous RNA segment to the bone marrow cells. By contacting, it is meant bringing the recombinant alphavirus and the bone marrow cells in physical proximity. The contacting step can be performed *in vitro* or *in vivo*. *In vitro* contacting can be carried out with cultures of immortalized or non-immortalized bone marrow cells. In one particular embodiment, bone marrow cells can be removed from a subject, cultured *in vitro*,  
25

-17-

infected with the vector, and then introduced back into the subject. Contacting is performed *in vivo* when the recombinant alphavirus is administered to a subject. Pharmaceutical formulations of recombinant alphavirus can be administered to a subject parenterally (*e.g.*, subcutaneous, intracerebral, intradermal, intramuscular, 5 intravenous and intraarticular) administration. Alternatively, pharmaceutical formulations of the present invention may be suitable for administration to the mucus membranes of a subject (*e.g.*, intranasal administration, by use of a dropper, swab, or inhaler). Methods of preparing infectious virus particles and pharmaceutical formulations thereof are discussed in more detail hereinbelow.

10 By "introducing" the heterologous RNA segment into the bone marrow cells it is meant infecting the bone marrow cells with recombinant alphavirus containing the heterologous RNA, such that the viral vector carrying the heterologous RNA enters the bone marrow cells and can be expressed therein. As used with respect to the present invention, when the heterologous RNA is 15 "expressed," it is meant that the heterologous RNA is transcribed. In particular embodiments of the invention in which it is desired to produce a protein or peptide, expression further includes the steps of post-transcriptional processing and translation of the mRNA transcribed from the heterologous RNA. In contrast, where the heterologous RNA encodes an antisense oligonucleotide, expression need 20 not include post-transcriptional processing and translation. With respect to embodiments in which the heterologous RNA encodes an immunogenic protein or a protein being administered for therapeutic purposes, expression may also include the further step of post-translational processing to produce an immunogenic or therapeutically-active protein.

25 The present invention also provides infectious RNAs, as described hereinabove, and cDNAs encoding the same. Preferably the infectious RNAs and cDNAs are derived from the S.A.AR86, Girdwood S.A., TR339, or Ockelbo viruses. The cDNA clones can be generated by any of a variety of suitable methods known to those skilled in the art. A preferred method is the method set

-18-

forth in United States Patent No. 5,185,440 to Davis et al., the disclosure of which is incorporated in its entirety by reference, and Gubler et al., *Gene* 25:263 (1983).

5 RNA is preferably synthesized from the DNA sequence *in vitro* using purified RNA polymerase in the presence of ribonucleotide triphosphates and cap analogs in accordance with conventional techniques. However, the RNA may also be synthesized intracellularly after introduction of the cDNA.

#### A. Double Promoter Vectors.

10 In one embodiment of the invention, double promoter vectors are used to introduce the heterologous RNA into the target bone marrow cells. A double promoter virus vector is a replication and propagation competent virus. Double promoter vectors are described in United States Patent No. 5,505,947 to Johnston et al., the disclosure of which is incorporated in its entirety by reference. Preferred alphaviruses for constructing the double promoter vectors are S.A.AR86, Girdwood S.A., TR339 and Ockelbo viruses. More preferably, the double 15 promoter vector contains one or more attenuating mutations. Attenuating mutations are described in more detail hereinabove.

20 The double promoter vector is constructed so as to contain a second subgenomic promoter (*i.e.*, 26S promoter) inserted 3' to the virus RNA encoding the structural proteins. The heterologous RNA is inserted between the second subgenomic promoter, so as to be operatively associated therewith, and the 3' UTR of the virus genome. Heterologous RNA sequences of less than 3 kilobases, more preferably those less than 2 kilobases, and more preferably still those less than 1 kilobase, can be inserted into the double promoter vector. In a preferred embodiment of the invention, the double promoter vector is derived from 25 Girdwood S.A., and the second subgenomic promoter is a duplicate of the Girdwood S.A. subgenomic promoter. In an alternate preferred embodiment, the double promoter vector is derived from TR339, and the second subgenomic promoter is a duplicate of the TR339 subgenomic promoter.

-19-

**B. Replicon Vectors.**

5           Replicon vectors, which are propagation-defective virus vectors can also be used to carry out the present invention. Replicon vectors are described in more detail in copending United States Application 08/448,630 to Johnston et al., the disclosure of which is incorporated in its entirety by reference. Preferred alphaviruses for constructing the replicon vectors are S.A.AR86, Girdwood S.A., TR339, and Ockelbo.

10           In general, in the replicon system, a foreign gene to be expressed is inserted in place of at least one of the viral structural protein genes in a transcription plasmid containing an otherwise full-length cDNA copy of the alphavirus genome RNA. RNA transcribed from this plasmid contains an intact copy of the viral nonstructural genes which are responsible for RNA replication and transcription. Thus, if the transcribed RNA is transfected into susceptible cells, it will be replicated and translated to give the nonstructural proteins. These 15           proteins will transcribe the transfected RNA to give high levels of subgenomic mRNA, which will then be translated to produce high levels of the foreign protein. The autonomously replicating RNA (*i.e.*, replicon) can only be packaged into virus particles if the alphavirus structural protein genes are provided on one or more "helper" RNAs, which are cotransfected into cells along with the replicon RNA. 20           The helper RNAs do not contain the viral nonstructural genes for replication, but these functions are provided *in trans* by the replicon RNA. Similarly, the transcriptase functions translated from the replicon RNA transcribe the structural protein genes on the helper RNA, resulting in the synthesis of viral structural proteins and packaging of the replicon into virus-like particles. As the packaging 25           or encapsidation signal for alphavirus RNAs is located within the nonstructural genes, the absence of these sequences in the helper RNAs precludes their incorporation into virus particles.

30           Alphavirus-permissive cells employed in the methods of the present invention are cells which, upon transfection with the viral RNA transcript, are capable of producing viral particles. Preferred alphavirus-permissive cells are

-20-

TR339-permissive cells, Girdwood S.A.-permissive cells, S.A.AR86-permissive cells, and Ockelbo-permissive cells. Alphaviruses have a broad host range. Examples of suitable host cells include, but are not limited to Vero cells, baby hamster kidney (BHK) cells, and chicken embryo fibroblast cells.

5       The phrase "structural protein" as used herein refers to the encoded proteins which are required for encapsidation (*e.g.*, packaging) of the RNA replicon, and include the capsid protein, E1 glycoprotein, and E2 glycoprotein. As described hereinabove, the structural proteins of the alphavirus are distributed among one or more helper RNAs (*i.e.*, a first helper RNA and a second helper RNA). In addition, one or  
10 more structural proteins may be located on the same RNA molecule as the replicon RNA, provided that at least one structural protein is deleted from the replicon RNA such that the resulting alphavirus particle is propagation defective. As used herein, the terms "deleted" or "deletion" mean either total deletion of the specified segment or the deletion of a sufficient portion of the specified segment to render the segment inoperative or  
15 nonfunctional, in accordance with standard usage. *See, e.g.*, U.S. Patent No. 4,650,764 to Temin et al. The term "propagation defective" as used herein, means that the replicon RNA cannot be encapsidated in the host cell in the absence of the helper RNA. The resulting alphavirus replicon particles are propagation defective inasmuch as the replicon RNA in these particles does not include all of the alphavirus structural proteins required  
20 for encapsidation, at least one of the required structural proteins being deleted therefrom, such that the replicon RNA initiates only an abortive infection; no new viral particles are produced, and there is no spread of the infection to other cells.

The helper cell for expressing the infectious, propagation defective alphavirus particle comprises a set of RNAs, as described above. The set of RNAs principally  
25 include a first helper RNA and a second helper RNA. The first helper RNA includes RNA encoding at least one alphavirus structural protein but does not encode all alphavirus structural proteins. In other words, the first helper RNA does not encode at least one alphavirus structural protein; the at least one non-coded alphavirus structural protein being deleted from the first helper RNA.

-21-

In one embodiment, the first helper RNA includes RNA encoding the alphavirus E1 glycoprotein, with the alphavirus capsid protein and the alphavirus E2 glycoprotein being deleted from the first helper RNA. In another embodiment, the first helper RNA includes RNA encoding the alphavirus E2 glycoprotein, with the alphavirus capsid protein and the alphavirus E1 glycoprotein being deleted from the first helper RNA. In a third, preferred embodiment, the first helper RNA includes RNA encoding the alphavirus E1 glycoprotein and the alphavirus E2 glycoprotein, with the alphavirus capsid protein being deleted from the first helper RNA.

The second helper RNA includes RNA encoding at least one alphavirus structural protein which is different from the at least one structural protein encoded by the first helper RNA. Thus, the second helper RNA encodes at least one alphavirus structural protein which is not encoded by the first helper RNA. The second helper RNA does not encode the at least one alphavirus structural protein which is encoded by the first helper RNA, thus the first and second helper RNAs do not encode duplicate structural proteins. In the embodiment wherein the first helper RNA includes RNA encoding only the alphavirus E1 glycoprotein, the second helper RNA may include RNA encoding one or both of the alphavirus capsid protein and the alphavirus E2 glycoprotein which are deleted from the first helper RNA. In the embodiment wherein, the first helper RNA includes RNA encoding only the alphavirus E2 glycoprotein, the second helper RNA may include RNA encoding one or both of the alphavirus capsid protein and the alphavirus E1 glycoprotein which are deleted from the first helper RNA. In the embodiment wherein the first helper RNA includes RNA encoding both the alphavirus E1 glycoprotein and the alphavirus E2 glycoprotein, the second helper RNA may include RNA encoding the alphavirus capsid protein which is deleted from the first helper RNA.

In one embodiment, the packaging segment (RNA comprising the encapsidation or packaging signal) is deleted from at least the first helper RNA.

-22-

In a preferred embodiment, the packaging segment is deleted from both the first helper RNA and the second helper RNA.

5           In the preferred embodiment wherein the packaging segment is deleted from both the first helper RNA and the second helper RNA, the helper cell is co-transfected with a replicon RNA in addition to the first helper RNA and the second helper RNA. The replicon RNA encodes the packaging segment and an inserted heterologous RNA. The inserted heterologous RNA may be RNA encoding a protein or a peptide. In a preferred embodiment, the replicon RNA, the first helper RNA and the second helper RNA are provided on separate molecules such that a first molecule, *i.e.*, the replicon RNA, includes RNA encoding the packaging segment and the inserted heterologous RNA, a second molecule, *i.e.*, the first helper RNA, includes RNA encoding at least one but not all of the required alphavirus structural proteins, and a third molecule, *i.e.*, the second helper RNA, includes RNA encoding at least one but not all of the required alphavirus structural proteins. For example, in one preferred embodiment of the present invention, the helper cell includes a set of RNAs which include (a) a replicon RNA including RNA encoding an alphavirus packaging sequence and an inserted heterologous RNA, (b) a first helper RNA including RNA encoding the alphavirus E1 glycoprotein and the alphavirus E2 glycoprotein, and (c) a second helper RNA including RNA encoding the alphavirus capsid protein so that the alphavirus E1 glycoprotein, the alphavirus E2 glycoprotein and the capsid protein assemble together into alphavirus particles in the host cell.

10

15

20

25           In an alternate embodiment, the replicon RNA and the first helper RNA are on separate molecules, and the replicon RNA and RNA encoding a structural gene not encoded by the first helper RNA are on another single molecule together, such that a first molecule, *i.e.*, the first helper RNA, including RNA encoding at least one but not all of the required alphavirus structural proteins, and a second molecule, *i.e.*, the replicon RNA, including RNA encoding the packaging segment, the inserted heterologous RNA, and the remaining structural proteins not encoded by the first helper RNA. For example, in one preferred embodiment of

30

-23-

the present invention, the helper cell includes a set of RNAs including (a) a replicon RNA including RNA encoding an alphavirus packaging sequence, an inserted heterologous RNA, and an alphavirus capsid protein, and (b) a first helper RNA including RNA encoding the alphavirus E1 glycoprotein and the alphavirus E2 glycoprotein so that the alphavirus E1 glycoprotein, the alphavirus E2 glycoprotein and the capsid protein assemble together into alphavirus particles in the host cell, with the replicon RNA packaged therein.

In one preferred embodiment of the present invention, the RNA encoding the alphavirus structural proteins, *i.e.*, the capsid, E1 glycoprotein and E2 glycoprotein, contains at least one attenuating mutation, as described hereinabove. Thus, according to this embodiment, at least one of the first helper RNA and the second helper RNA includes at least one attenuating mutation. In a more preferred embodiment, at least one of the first helper RNA and the second helper RNA includes at least two, or multiple, attenuating mutations. The multiple attenuating mutations may be positioned in either the first helper RNA or in the second helper RNA, or they may be distributed randomly with one or more attenuating mutations being positioned in the first helper RNA and one or more attenuating mutations positioned in the second helper RNA. Alternatively, when the replicon RNA and the RNA encoding the structural proteins not encoded by the first helper RNA are located on the same molecule, an attenuating mutation may be positioned in the RNA which codes for the structural protein not encoded by the first helper RNA. The attenuating mutations may also be located within the RNA encoding non-structural proteins (*e.g.*, the replicon RNA).

Preferably, the first helper RNA and the second helper RNA also include a promoter. It is also preferred that the replicon RNA also includes a promoter. Suitable promoters for inclusion in the first helper RNA, second helper RNA and replicon RNA are well known in the art. One preferred promoter is the Girdwood S.A. 26S promoter for use when the alphavirus is Girdwood S.A. Another preferred promoter is the TR339 26S promoter for use when the alphavirus is TR339. Additional promoters beyond the Girdwood S.A. and TR339

-24-

promoters include the VEE 26S promoter, the Sindbis 26S promoter, the Semliki Forest 26S promoter, and any other promoter sequence recognized by alphavirus polymerases. Alphavirus promoter sequences containing mutations which alter the activity level of the promoter (in relation to the activity level of the wild-type) are also suitable in the practice of the present invention. Such mutant promoter sequences are described in Raju and Huang, *J. Virol.* 65, 2501-2510 (1991), the disclosure of which is incorporated herein in its entirety. In the system wherein the first helper RNA, the second helper RNA, and the replicon RNA are all on separate molecules, the promoters, if the same promoter is used for all three RNAs, provide a homologous sequence between the three molecules. It is preferred that the selected promoter is operative with the non-structural proteins encoded by the replicon RNA molecule.

In cases where vaccination with two immunogens provides improved protection against disease as compared to vaccination with only a single immunogen, a double-promoter replicon would ensure that both immunogens are produced in the same cell. Such a replicon would be the same as the one described above, except that it would contain two copies of the 26S RNA promoter, each followed by a different multiple cloning site, to allow for the insertion and expression of two different heterologous proteins. Another useful strategy is to insert the IRES sequence from the picornavirus, EMC virus, between the two heterologous genes downstream from the single 26S promoter of the replicon described above, thus leading to expression of two immunogens from the single replicon transcript in the same cell.

C. Uses of the Present Invention.

The alphavirus vectors, RNAs, cDNAs, helper cells, infectious virus particles, and methods of the present invention find use in *in vitro* expression systems, wherein the inserted heterologous RNA encodes a protein or peptide which is desirably produced *in vitro*. The RNAs, cDNAs, helper cells, infectious virus particles, methods, and pharmaceutical formulations of the present invention are additionally useful in a method of administering a protein or peptide to a

-25-

subject in need of the protein or peptide, as a method of treatment or otherwise. In this embodiment of the invention, the heterologous RNA encodes the desired protein or peptide, and pharmaceutical formulations of the present invention are administered to a subject in need of the desired protein or peptide. In this manner, 5 the protein or peptide may thus be produced *in vivo* in the subject. The subject may be in need of the protein or peptide because the subject has a deficiency thereof, or because the production of the protein or peptide in the subject may impart some therapeutic effect, as a method of treatment or otherwise.

10 Alternately, the claimed methods provide a vaccination strategy, wherein the heterologous RNA encodes an immunogenic protein or peptide.

The methods and products of the invention are also useful as antigens and for evoking the production of antibodies in animals such as horses and rabbits, from which the antibodies may be collected and then used in diagnostic assays in accordance with known techniques.

15 A further aspect of the present invention is a method of introducing and expressing antisense oligonucleotides in bone marrow cell cultures to regulate gene expression. Alternately, the claimed method finds use in introducing and expressing a protein or peptide in bone marrow cell cultures.

II. Girdwood S.A. and TR339 Clones.

20 Disclosed hereinbelow are genomic RNA sequences encoding live Girdwood S.A. virus, live S.A.AR86 virus, and live Sindbis strain TR339 virus, cDNAs derived therefrom, infectious RNA transcripts encoded by the cDNAs, infectious viral particles containing the infectious RNA transcripts, and pharmaceutical formulations derived therefrom.

25 The cDNA sequence of Girdwood S.A. is given herein as SEQ ID NO:4. Alternatively, the cDNA may have a sequence which differs from the cDNA of SEQ ID NO:4, but which has the same protein sequence as the cDNA

-26-

given herein as SEQ ID NO:4. Thus, the cDNA may include one or more silent mutations.

5                 The phrase "silent mutation" as used herein refers to mutations in the cDNA coding sequence which do not produce mutations in the corresponding protein sequence translated therefrom.

10                Likewise, the cDNA sequence of TR339 is given herein as SEQ ID NO:8. Alternatively, the cDNA may have a sequence which differs from the cDNA of SEQ ID NO:8, but which has the same protein sequence as the cDNA given herein as SEQ ID NO:8. Thus, the cDNA may include one or more silent mutations.

15                The cDNAs encoding infectious Girdwood S.A. and TR339 virus RNA transcripts of the present invention include those homologous to, and having essentially the same biological properties as, the cDNA sequences disclosed herein as SEQ ID NO:4 and SEQ ID NO:8, respectively. Thus, cDNAs that hybridize to cDNAs encoding infectious Girdwood S.A. or TR339 virus RNA transcripts disclosed herein are also an aspect of this invention. Conditions which will permit other cDNAs encoding infectious Girdwood S.A. or TR339 virus transcripts to hybridize to the cDNAs disclosed herein can be determined in accordance with known techniques. For example, hybridization of such sequences may be carried 20                out under conditions of reduced stringency, medium stringency, or even high stringency conditions (e.g., conditions represented by a wash stringency of 35-40% formamide with 5X Denhardt's solution, 0.5% SDS and 1X SSPE at 37°C; conditions represented by a wash stringency of 40-45% formamide with 5X Denhardt's solution, 0.5% SDS, and 1X SSPE at 42°C; and conditions represented 25                by a wash stringency of 50% formamide with 5X Denhardt's solution, 0.5% SDS and 1X SSPE at 42°C, respectively, to cDNA encoding infectious Girdwood S.A. or TR339 virus RNA transcripts disclosed herein in a standard hybridization assay. See J. SAMBROOK ET AL., MOLECULAR CLONING: A LABORATORY MANUAL (2d ed. 1989)). In general, cDNA sequences encoding infectious

-27-

Girdwood S.A. or TR339 virus RNA transcripts that hybridize to the cDNAs disclosed herein will be at least 30% homologous, 50% homologous, 75% homologous, and even 95% homologous or more with the cDNA sequences encoding infectious Girdwood S.A. or TR339 virus RNA transcripts disclosed  
5 herein.

Promoter sequences and Girdwood S.A. virus or Sindbis virus strain TR339 cDNA clones are operatively associated in the present invention such that the promoter causes the cDNA clone to be transcribed in the presence of an RNA polymerase which binds to the promoter. The promoter is positioned on the 5' end  
10 (with respect to the virion RNA sequence), of the cDNA clone. An excessive number of nucleotides between the promoter sequence and the cDNA clone will result in the inoperability of the construct. Hence, the number of nucleotides between the promoter sequence and the cDNA clone is preferably not more than eight, more preferably not more than five, still more preferably not more than  
15 three, and most preferably not more than one.

Examples of promoters which are useful in the cDNA sequences of the present invention include, but are not limited to T3 promoters, T7 promoters, cytomegalovirus (CMV) promoters, and SP6 promoters. The DNA sequence of the present invention may reside in any suitable transcription vector. The DNA sequence preferably has a complementary DNA sequence bound thereto so that the double-stranded sequence will serve as an active template for RNA polymerase.  
20 The transcription vector preferably comprises a plasmid. When the DNA sequence comprises a plasmid, it is preferred that a unique restriction site be provided 3' (with respect to the virion RNA sequence) to the cDNA clone. This provides a means for linearizing the DNA sequence to allow the transcription of genome-length RNA *in vitro*.  
25

The cDNA clones can be generated by any of a variety of suitable methods known to those skilled in the art. A preferred method is the method set forth in United States Patent No. 5,185,440 to Davis et al., the disclosure of which

-28-

is incorporated in its entirety by reference, and Gubler et al., *Gene* 25:263 (1983).

RNA is preferably synthesized from the DNA sequence *in vitro* using purified RNA polymerase in the presence of ribonucleotide triphosphates and cap analogs in accordance with conventional techniques. However, the RNA may 5 also be synthesized intracellularly after introduction of the cDNA.

The Girdwood S.A. and TR339 cDNA clones and the infectious RNAs and infectious virus particles produced therefrom of the present invention are useful for the preparation of pharmaceutical formulations, such as vaccines. In addition, the cDNA clones, infectious RNAs, and infectious viral particles of 10 the present invention are useful for administration to animals for the purpose of producing antibodies to the Girdwood S.A. virus or the Sindbis virus strain TR339, which antibodies may be collected and used in known diagnostic techniques for the detection of Girdwood S.A. virus or Sindbis virus strain TR339. Antibodies can also be generated to the viral proteins expressed from the cDNAs 15 disclosed herein. As another aspect of the present invention, the claimed cDNA clones are useful as nucleotide probes to detect the presence of Girdwood S.A. or TR339 genomic RNA or transcripts.

### III. Infectious Virus Particles and Pharmaceutical Formulations.

The infectious virus particles of the present invention include those containing double promoter vectors and those containing replicon vectors as 20 described hereinabove. Alternately, the infectious virus particles contain infectious RNAs encoding the Girdwood S.A. or TR339 genome. When the infectious RNA comprises the Girdwood S.A. genome, preferably the RNA has the sequence encoded by the cDNA given as SEQ ID NO:4. When the infectious RNA 25 comprises the TR339 genome, preferably the RNA has the sequence encoded by the cDNA given as SEQ ID NO:8.

The infectious, alphavirus particles of the present invention may be prepared according to the methods disclosed herein in combination with techniques

-29-

known to those skilled in the art. These methods include transfecting an alphavirus-permissive cell with a replicon RNA including the alphavirus packaging segment and an inserted heterologous RNA, a first helper RNA including RNA encoding at least one alphavirus structural protein, and a second helper RNA including RNA encoding at least one alphavirus structural protein which is different from that encoded by the first helper RNA. Alternately, and preferably, at least one of the helper RNAs is produced from a cDNA encoding the helper RNA and operably associated with an appropriate promoter, the cDNA being stably transfected and integrated into the cells. More preferably, all of the helper RNAs will be "launched" from stably transfected cDNAs. The step of transfecting the alphavirus-permissive cell can be carried out according to any suitable means known to those skilled in the art, as described above with respect to propagation-competent viruses.

Uptake of propagation-competent RNA into the cells *in vitro* can be carried out according to any suitable means known to those skilled in the art. Uptake of RNA into the cells can be achieved, for example, by treating the cells with DEAE-dextran, treating the RNA with LIPOFECTIN® before addition to the cells, or by electroporation, with electroporation being the currently preferred means. These techniques are well known in the art. See e.g., United States Patent No. 5,185,440 to Davis et al., and PCT Publication No. WO 92/10578 to Biooption AB, the disclosures of which are incorporated herein by reference in their entirety. Uptake of propagation-competent RNA into the cell *in vivo* can be carried out by administering the infectious RNA to a subject as described in Section I above.

The infectious RNAs may also contain a heterologous RNA segment, where the heterologous RNA segment contains a heterologous RNA and a promoter operably associated therewith. It is preferred that the infectious RNA introduces and expresses the heterologous RNA in bone marrow cells as described in Section I above. According to this embodiment, it is preferable that the promoter operatively associated with the heterologous RNA is operable in bone

-30-

5 marrow cells. The heterologous RNA may encode any protein or peptide, preferably an immunogenic protein or peptide, a therapeutic protein or peptide, a hormone, a growth factor, an interleukin, a cytokine, a chemokine, an enzyme, a ribozyme, or an antisense oligonucleotide as described in more detail in Section I above.

10 The step of facilitating the production of the infectious viral particles in the cells may be carried out using conventional techniques. *See e.g.*, United States Patent No. 5,185,440 to Davis et al., PCT Publication No. WO 92/10578 to Bioption AB, and United States Patent No. 4,650,764 to Temin et al. (although Temin et al., relates to retroviruses rather than alphaviruses).

15 The step of collecting the infectious virus particles may also be carried out using conventional techniques. For example, the infectious particles may be collected by cell lysis, or collection of the supernatant of the cell culture, as is known in the art. *See e.g.*, United States Patent No. 5,185,440 to Davis et al., PCT Publication No. WO 92/10578 to Bioption AB, and United States Patent No. 4,650,764 to Temin et al. Other suitable techniques will be known to those skilled in the art. Optionally, the collected infectious virus particles may be purified if desired. Suitable purification techniques are well known to those skilled 20 in the art.

25 Pharmaceutical formulations, such as vaccines, of the present invention comprise an immunogenic amount of the infectious, virus particles in combination with a pharmaceutically acceptable carrier. An "immunogenic amount" is an amount of the infectious virus particles which is sufficient to evoke an immune response in the subject to which the pharmaceutical formulation is administered. An amount of from about  $10^3$  to about  $10^7$  particles, and preferably about  $10^4$  to  $10^6$  particles per dose is believed suitable, depending upon the age and species of the subject being treated, and the immunogen against which the immune response is desired.

-31-

Pharmaceutical formulations of the present invention for therapeutic use comprise a therapeutic amount of the infectious virus particles in combination with a pharmaceutically acceptable carrier. A "therapeutic amount" is an amount of the infectious virus particles which is sufficient to produce a therapeutic effect (e.g., triggering an immune response or supplying a protein to a subject in need thereof) in the subject to which the pharmaceutical formulation is administered. The therapeutic amount will depend upon the age and species of the subject being treated, and the therapeutic protein or peptide being administered. Typical dosages are an amount from about 10<sup>1</sup> to about 10<sup>5</sup> infectious units.

Exemplary pharmaceutically acceptable carriers include, but are not limited to, sterile pyrogen-free water and sterile pyrogen-free physiological saline solution. Subjects which may be administered immunogenic amounts of the infectious virus particles of the present invention include but are not limited to human and animal (e.g., pig, cattle, dog, horse, donkey, mouse, hamster, monkeys) subjects.

Pharmaceutical formulations of the present invention include those suitable for parenteral (e.g., subcutaneous, intracerebral, intradermal, intramuscular, intravenous and intraarticular) administration. Alternatively, pharmaceutical formulations of the present invention may be suitable for administration to the mucus membranes of a subject (e.g., intranasal administration by use of a dropper, swab, or inhaler). The formulations may be conveniently prepared in unit dosage form and may be prepared by any of the methods well known in the art.

The following examples are provided to illustrate the present invention, and should not be construed as limiting thereof. In these examples, PBS means phosphate buffered saline, EDTA means ethylene diamine tetraacetate, ml means milliliter,  $\mu$ l means microliter, mM means millimolar,  $\mu$ M means micromolar, u means unit, PFU means plaque forming units, g means gram, mg means milligram,  $\mu$ g means microgram, cpm means counts per minute, ic means

-32-

intracerebral or intracerebrally, ip means intraperitoneal or intraperitoneally, iv means intravenous or intravenously, and sc means subcutaneous or subcutaneously.

Amino acid sequences disclosed herein are presented in the amino to carboxyl direction, from left to right. The amino and carboxyl groups are not presented in the sequence. Nucleotide sequences are presented herein by single strand only in the 5' to 3' direction, from left to right. Nucleotides and amino acids are represented herein in the manner recommended by the IUPAC-IUB Biochemical Nomenclature Commission, or (for amino acids) by either one letter or three letter code, in accordance with 37 CFR § 1.82<sup>2</sup> and established usage.  
5 Where one letter amino acid code is used, the same sequence is also presented elsewhere in three letter code.  
10

#### EXAMPLE I

##### Cells and Virus Stocks

S.A.AR86 was isolated in 1954 from a pool of *Culex* sp. mosquitoes collected near Johannesburg, South Africa. Weinbren et al., *S. Afr. Med. J.* 30, 631-36 (1956). Ockelbo82 was isolated from *Culiseta* sp. mosquitoes collected in Edsbyn, Sweden in 1982 and was associated serologically with human disease. Niklasson et al., *Am. J. Trop. Med. Hyg.* 33, 1212-17 (1984). Girdwood S.A. was isolated from a human patient in the Johannesburg area of South Africa in 1963. Malherbe et al., *S. Afr. Med. J.* 37, 547-52 (1963). Molecularly cloned virus TR339 represents the deduced consensus sequence of Sindbis AR339. McKnight et al., *J. Virol.* 70, 1981-89 (1996); William Klimstra, personal communication. TRSB is a laboratory strain of Sindbis isolate AR339 derived from a cDNA clone pTRSB and differing from the AR339 consensus sequence at 20 three codons. McKnight et al., *J. Virol.* 70, 1981-89 (1996). pTRS000 is a full-length cDNA clone of Sindbis AR339 following the SP6 phage promoter and containing mostly Sindbis AR339 sequences.  
25

Stocks of all molecularly cloned viruses were prepared by electroporating genome length *in vitro* transcripts of their respective cDNA clones

-33-

in BHK-21 cells. Heidner et al., *J. Virol.* 68, 2683-92 (1994). Girdwood S.A. (Malherbe et al., *S. Afr. Med. J.* 37, 547-52 (1963)) and Ockelbo82 (Espmark and Niklasson, *Am. J. Trop. Med. Hyg.* 33, 1203-11 (1984); Niklasson et al., *Am. J. Trop. Med. Hyg.* 33, 1212-17 (1984)) were passed one to three times in BHK-21  
5 cells in order to produce amplified stocks of virus. All virus stocks were stored at -70°C until needed. The titers of the virus stocks were determined on BHK-21 cells from aliquots of frozen virus.

#### EXAMPLE 2

##### Cloning the S.A.AR86 and Girdwood S.A. Genomic Sequences

10 The sequences of S.A.AR86 (Figure 1, SEQ ID NO: 1) and Girdwood S.A. (Figure 3, SEQ ID NO:4) were determined from uncloned reverse transcriptase-polymerase chain reaction (RT-PCR) fragments amplified from virion RNA. Heidner et al., *J. Virol.* 68, 2683-92 (1994). The sequence of the 5' 40 nucleotides was determined by directly sequencing the genomic RNA. Sanger et al., *Proc. Natl. Acad. Sci. USA* 74, 5463-67 (1977); Zimmern and Kaesberg, *Proc. Natl. Acad. Sci. USA* 75, 4257-61 (1978); Ahlquist et al., *Cell* 23, 183-89  
15 (1981).

20 The S.A.AR86 genome was 11,663 nucleotides in length, excluding the 5' CAP and 3'poly(A) tail, 40 nucleotides shorter than the alphavirus prototype Sindbis strain AR339. Strauss et al., *Virology* 133, 92-110 (1984). Compared with the consensus sequence of Sindbis virus AR339 (McKnight et al., *J. Virol.* 70 1981-89 (1996)), S.A.AR86 contained two separate 6-nucleotide insertions, and one 3-nucleotide insertion in the 3' half of the nsP3 gene, a region not well conserved among alphaviruses. The two 6-nucleotide insertions were found immediately 3' of nucleotides 5403 and 5450, and the 3-nucleotide insertion was immediately 3' of nucleotide 5546 compared with the AR339 genome. In addition, S.A.AR86 contained a 54-nucleotide deletion in nsP3 which spanned nucleotides 5256 to 5311 of AR339. As a result of these deletions and insertions, S.A.AR86 nsP3 was 13 amino acids smaller than AR339, containing an 18-amino acid deletion and a total of 5 amino acids inserted. The 3' untranslated region of  
25  
30

-34-

S.A.AR86 contained, with respect to AR339, two 1-nucleotide deletions at nucleotides 11,513 and 11,602, and one 1-nucleotide insertion following nucleotide 11,664. The total numbers of nucleotides and predicted amino acids comprising the remaining genes of S.A.AR86 were identical to those of AR339.

5           A notable feature of the deduced amino acid sequence of S.A.AR86 (Figure 2, SEQ ID NO:2 and SEQ ID NO:3) was the cysteine codon in place of an opal termination codon between nsP3 and nsP4. S.A.AR86 is the only alphavirus of the Sindbis group, and one of just three alphavirus isolates sequenced to date, which do not contain an opal termination codon between nsP3 and nsP4.  
10           Takkinen, K., *Nucleic Acids Res.* 14, 5667-5682 (1986); Strauss et al., *Virology* 164, 265-74 (1988).

15           The genome of Girdwood S.A. was 11,717 nucleotides long excluding the 5' CAP and 3' poly(A) tail. The nucleotide sequence (SEQ ID NO:4) of the Girdwood S.A. genome and the putative amino acid sequence (SEQ ID NO:5 and SEQ ID NO:6) of the Girdwood S.A. gene products are shown in Figure 3 and Figure 4, respectively. The asterisk at position 1902 in SEQ ID NO:5 indicates the position of the opal termination codon in the coding region of the nonstructural polyprotein. The extra nucleotides relative to AR339 were in the nonconserved half of nsP3, which contained insertions totalling 15 nucleotides, and in the 3' untranslated region which contained two 1-nucleotide deletions and a 1-nucleotide insertion with respect to the consensus Sindbis AR339 genome. The insertions found in the nsP3 gene of Girdwood S.A. were identical in position and content to those found in S.A.AR86, although Girdwood S.A. did not have the large nsP3 deletion characteristic of S.A.AR86. The remaining portions of the genome contained the same number of nucleotides and predicted amino acids as Sindbis AR339.  
20  
25

Overall, Girdwood S.A. was 94.5% identical to the consensus Sindbis AR339 sequence, differing at 655 nucleotides not including the insertions and deletions. These nucleotide differences resulted in 88 predicted amino acid

-35-

changes or a difference of 2.3%. A plurality of amino acid differences were concentrated in the nsP3 gene, which contained 32 of the amino acid changes, 25 of which were in the nonconserved 3' half.

5           The Girdwood S.A. nucleotides at positions 1, 3, and 11,717 could not be resolved. Because the primer used during the RT-PCR amplification of the 3' end of the genome assumed a cytosine in the 3' terminal position, the identity of this nucleotide could not be determined with certainty. However, in all alphaviruses sequenced to date there is a cytosine in this position. This, combined with the fact that no difficulty was encountered in obtaining RT-PCR product for 10 this region with an oligo(dT) primer ending with a 3'G, suggested that Girdwood S.A. also contains a cytosine at this position. The ambiguity at nucleotide positions 1 and 3 resulted from strong stops encountered during the RNA sequencing.

### EXAMPLE 3

15           Comparison of S.A.AR86 and Girdwood S.A.  
Sequences With Other Sindbis-Related Virus Sequences

20           Table 1 examines the relationship of S.A.AR86 and Girdwood S.A. to each other and to other Sindbis-related viruses. This was accomplished by aligning the nucleotide and deduced amino acid sequences of Ockelbo82, AR339 and Girdwood S.A. to those of S.A.AR86 and then calculating the percentage identity for each gene using the programs contained within the Wisconsin GCG package (Genetics Computer Group, 575 Science Drive, Madison WI 53711); as described in more detail in McKnight et al., *J. Virol.* 70, 1981-89 (1996).

25           The analysis suggests that S.A.AR86 is most similar to the other South African isolate, Girdwood S.A., and that the South African isolates are more similar to the Swedish Ockelbo82 isolate than to the Egyptian Sindbis AR339 isolate. These results also suggest that it is unlikely that S.A.AR86 is a recombinant virus like WEE virus. Hahn et al., *Proc. Natl. Acad. Sci. USA* 85, 5997-6001 (1988).

-36-

**TABLE I**  
**Comparison of the Nucleotide and Amino Acid Sequences**  
**of S.A.AR86 Virus with Those of Sindbis AR339, OCK82, and Girdwood S.A. Viruses<sup>a</sup>**

Regions	Nucleotide Differences <sup>b</sup>				Amino Acid Differences <sup>b</sup>	
	AR339	OCK82	GIRD	AR339	OCK82	GIRD
5' untranslated	0 (0.0)	0 (0.0)	1 (1.7)	--	--	--
nsP1	76 (4.7)	37 (2.3)	15 (0.9)	9 (1.7)	6 (1.1)	2 (0.4)
nsP2	137 (5.7)	86 (3.6)	45 (1.9)	15 (1.9)	8 (1.0)	12 (1.5)
nsP3	51 (5.7)	35 (3.9)	13 (1.6)	6 (2.0)	1 (0.3)	1 (0.4)
Conserved <sup>c</sup>	116 (6.6)	83 (4.4)	70 (2.2)	45 (9.7)	34 (7.0)	27 (3.7)
Nonconserved <sup>d</sup>	111 (6.1)	68 (3.7)	19 (1.1)	8 (1.3)	2 (0.3)	4 (0.6)
26s junction	1 (2.1)	0 (0.0)	1 (2.1)	--	--	--
Capsid	36 (4.5)	26 (3.3)	7 (0.9)	1 (0.4)	3 (1.1)	0 (0.0)
E3	17 (8.9)	5 (2.6)	4 (2.1)	1 (1.6)	0 (0.0)	0 (0.0)
E2	71 (5.6)	43 (3.4)	18 (1.4)	12 (2.6)	6 (1.4)	2 (0.5)
6K	10 (6.1)	9 (5.4)	4 (2.4)	2 (3.6)	2 (3.6)	1 (1.8)
E1	49 (3.7)	31 (2.3)	16 (1.2)	7 (1.6)	6 (1.4)	2 (0.9)
3' untranslated	14 (4.5)	8 (2.5)	1 (0.3)	--	--	--
<b>Totals</b>	<b>689 (5.5)</b>	<b>431 (3.3)</b>	<b>214 (1.4)</b>	<b>106 (2.3)</b>	<b>68 (1.4)</b>	<b>51 (0.9)</b>

<sup>a</sup> All nucleotide positions and gene boundaries are numbered according to those used for the Sindbis AR339, HR<sub>sp</sub> variant Genbank Accession No. J02363; Strauss et al., *Virology* 133, 92-110 (1984).

<sup>b</sup> Differences include insertions and deletions.

<sup>c</sup> Conserved region nucleotides 4100 to 5000 (aa 1 to aa300).

<sup>d</sup> Nonconserved region nucleotides 5001 to 5729 (aa301 to aa542, S.A.AR86 numbering).

-37-

## EXAMPLE 4

Neurovirulence of S.A.AR86 and Girdwood S.A.

Girdwood S.A., Ockelbo82, and S.A.AR86 are related by sequence; in contrast, it has previously been reported that only S.A.AR86 displayed the adult mouse neurovirulence phenotype. Russell et al., *J. Virol.* 63, 1619-29 (1989). These findings were confirmed by the present investigations. Briefly, groups of four female CD-1 mice (3-6 weeks of age) were inoculated ic with 10<sup>3</sup> plaque-forming units (PFU) of S.A.AR86, Girdwood S.A., or Ockelbo82. Neither Girdwood S.A. nor Ockelbo82 infection produced any clinical signs of infection. Infection with S.A.AR86 produced neurological signs within four to five days and ultimately killed 100% of the mice as previously demonstrated.

Table 2 lists those amino acids of S.A.AR86 which might explain the neurovirulence phenotype in adult mice. A position was scored as potentially related to the S.A.AR86 adult neurovirulence phenotype if the S.A.AR86 amino acid differed from that which otherwise was absolutely conserved at that position in the other viruses.

TABLE 2  
Divergent Amino Acids in S.A.AR86  
Potentially Related to the Adult Neurovirulence Phenotype

		Position in S.A.AR86	S.A.AR86 Amino Acid	Conserved Amino Acid
20	nsP1	583	Thr	Ile
	nsP2	256	Arg	Ala
		648	Ile	Val
		651	Lys	Glu
25	nsP3	344	Gly	Glu
		386	Tyr	Ser
		441	Asp	Gly
		445	Ile	Met
		537	Cys	Opal
	E2	243	Ser	Leu
	6K	30	Val	Ile
	E1	112	Val	Ala
		169	Leu	Ser

-38-

EXAMPLE 5

pS55 Molecular Clone of S.A.AR86

As a first step in investigating the unique adult mouse neurovirulence phenotype of S.A.AR86, a full-length cDNA clone of the 5 S.A.AR86 genome was constructed. The sources of cDNA included conventional cDNA clones (Davis et al., *Virology* 171, 189-204 (1989)) as well as uncloned RT-PCR fragments derived from the S.A.AR86 genome. As described previously, these were substituted, starting at the 3' end, into pTR5000 (McKnight et al., *J. Virol.* 70, 1981-89 (1996)), a full-length Sindbis clone from which infectious 10 genomic replicas could be derived by transcription with SP6 polymerase *in vitro*.

The end result was pS55, a molecular clone of S.A.AR86 from 15 which infectious transcripts could be produced and which contained four nucleotide changes (G for A at nt 215; G for C at nt 3863; G for A at nt 5984; and C for T at nt 9113) but no amino acid coding differences with respect to the S.A.AR86 genomic RNA (amino acid sequence of S.A.AR86 presented in Figure 2 (SEQ ID NO:2 and SEQ ID NO:3)). The nucleotide sequence of clone pS55 is presented in Figure 5 (SEQ ID NO:7).

As has been described by Simpson et al., *Virology* 222, 464-69 20 (1996), neurovirulence and replication of the virus derived from pS55 (S55) were compared with those of S.A.AR86. It was found that S55 exhibits the distinctive adult neurovirulence characteristic of S.A.AR86. Like S.A.AR86, S55 produces 100% mortality in adult mice infected with the virus and the survival times of animals infected with both viruses were indistinguishable. In addition, S55 and 25 S.A.AR86 were found to replicate to essentially equivalent titers *in vivo*, and the profiles of S55 and S.A.AR86 virus growth in the central nervous system and periphery were very similar.

From these data it was concluded that the silent changes found in 30 virus derived from clone pS55 had little or no effect on its growth or virulence, and that this molecularly cloned virus accurately represents the biological isolate, S.A.AR86.

## EXAMPLE 6

Construction of the Consensus AR339 Virus TR339

The consensus sequence of the Sindbis virus AR339 isolate, the prototype alphavirus was deduced. The consensus AR339 sequence was inferred by comparison of the TRSB sequence (a laboratory-derived AR339 strain) with the complete or partial sequences of HR<sub>s</sub>p (the Gen Bank sequence; Strauss et al., *Virology* 133, 92-110 (1984)), SV1A, and NSV (AR339-derived laboratory strains; Lustig et al., *J. Virol.* 62, 2329-36 (1988)), and SIN (a laboratory-derived AR339 strain; Davis et al., *Virology* 161, 101-108 (1987), Strauss et al., *J. Virol.* 65, 10 4654-64 (1991)). Each of these viruses was descended from AR339. Where these sequences differed from each other, they also were compared with the amino acid sequences of other viruses related to Sindbis virus: Ockelbo82, S.A.AR86, Girdwood S.A., and the somewhat more distantly related Aura virus. Rumenapf et al., *Virology* 208, 621-33 (1995).

15 The details of determining a consensus AR339 sequence and constructing the consensus virus TR339 have been described elsewhere. McKnight et al., *J. Virol.* 70, 1981-89 (1996); Klimstra et al., *manuscript in preparation*. The nucleotide (SEQ ID NO:8) sequence of pTR339 is presented in Figure 6. The deduced amino acid sequences of the pTR339 non-structural and structural 20 polyproteins are shown as SEQ ID NO:9 and SEQ ID NO:10, respectively. The asterisk at position 1897 in SEQ ID NO:9 indicates the position of the opal termination codon in the coding region of the nonstructural polyprotein. The consensus nucleotide sequence diverged from the pTRSB sequence at three coding positions (nsP3 528, E2 1, and E1 72). These differences are illustrated in Table 25 3.

TABLE 3

Amino Acid Differences Between  
Laboratory Strain TRSB and Molecular Clone TR339

	nsP3 528 (nt5683)	E2 1 (nt8633)	E1 72 (nt10279)
TR339	Arg (CGA)	Ser (AGC)	Ala (GCU)
TRSB	Gln (CAA)	Arg (AGA)	Val (GUU)

**EXAMPLE 7****Animals Used for *In Vivo* Localization Studies**

Specific pathogen free CD-1 mice were obtained from Charles River Breeding Laboratories (Raleigh, North Carolina) at 21 days of age and maintained under barrier conditions until approximately 37 days of age. Intracerebral (ic) inoculations were performed as previously described, Simpson et al., *Virol.* 222, 464-49 (1996), with 500 PFU of S51 (an attenuated mutant of S55) or 10<sup>3</sup> PFU of S55. Animals inoculated peripherally were first anesthetized with METOFANE®. Then, 25 µl of diluent (PBS, pH 7.2, 1% donor calf serum, 100 u/ml penicillin, 10 50 µg/ml streptomycin, 0.9 mM CaCl<sub>2</sub>, and 0.5 mM MgCl<sub>2</sub>) containing 10<sup>3</sup> PFU of virus were injected either intravenously (iv) into the tail vein, subcutaneously (sc) into the skin above the shoulder blades on the middle of the back, or intraperitoneally (ip) in the lower right abdomen. Animals were sacrificed at various times post-inoculation as previously described. Simpson et al., *Virol.* 222, 464-49 (1996). Brains (including brainstems) were homogenized in diluent to 30% w/v, and right quadriceps were homogenized in diluent to 25% w/v. Homogenates were handled and titered as described previously. Simpson et al., *Virol.* 222, 464-49 (1996). Bone marrow was harvested by crushing both femurs from each animal in sufficient diluent to produce a 30% w/v suspension (calculated as weight of uncrushed femurs in volume of diluent). Samples were stored at -70°C. For titration, samples were thawed and clarified by centrifugation at 1,000 x g for 20 minutes at 4°C before being titered by conventional plaque assay on BHK-21 cells.

**EXAMPLE 8****Tissue Preparation for *In Situ* Hybridization Studies**

25 Animals were anesthetized by ip injection of 0.5 ml AVERTIN® at various times post-inoculation followed by perfusion with 60 to 75 ml of 4% paraformaldehyde in PBS (pH 7.2) at a flow rate of 10 ml per minute. The entire carcass was decalcified for 8 to 10 weeks in 4% paraformaldehyde containing 8% EDTA in PBS (pH 6.8) at 4°C. This solution was changed twice during the 30 decalcification period. Selected tissues were cut into blocks approximately 3 mm thick and placed into biopsy cassettes for paraffin embedding and sectioning. Blocks were embedded, sectioned and hematoxylin/eosin stained by Experimental Pathology Laboratories (Research Triangle Park, North Carolina) or North

Carolina State University Veterinary School Pathology Laboratory (Raleigh, North Carolina).

#### EXAMPLE 9

##### In Situ Hybridization

5        Hybridizations were performed using a [<sup>35</sup>S]-UTP labeled S.A.AR86 specific riboprobe derived from pDS-45. Clone pDS-45 was constructed by first amplifying a 707 base pair fragment from pS55 by PCR using primers 7241 (5'-CTGCGGCGATTCATCTTGC-3', SEQ ID NO:11) and SC-3 (5'-CTCCAACTTAAGTG-3', SEQ ID NO:12). The resulting 707 base pair fragment  
10      was purified using a GENE CLEAN® kit (Bio101, CA), digested with *Hha*I, and cloned into the *Sma*I site of pSP72 (Promega). Linearizing pDS-45 with *Eco*RV and performing an *in vitro* transcription reaction with SP6 DNA-dependent, RNA polymerase (Promega) in the presence of [<sup>35</sup>S]-UTP resulted in a riboprobe approximately 500 nucleotides in length of which 445 nucleotides were  
15      complementary to the S.A.AR86 genome (nucleotides 7371 through 7816). A riboprobe specific for the influenza strain PR-8 hemagglutinin (HA) gene was used as a control probe to test non-specific binding. The *in situ* hybridizations were performed as described previously (Charles et al., *Virol.* 208, 662-71 (1995)) using 10<sup>5</sup> cpm of probe per slide.

20

#### EXAMPLE 10

##### Replication of S.A.AR86 in Bone Marrow

25      Three groups of six adult mice each were inoculated peripherally (sc, ip, or iv) with 1200 PFU of S55 (a molecular clone of S.A.AR86) in 25 µl of diluent. Under these conditions, the infection produced no morbidity or mortality. Two mice from each group were anesthetized and sacrificed at 2, 4 and 6 days post-inoculation by exsanguination. The serum, brain (including brainstem), right quadricep, and both femurs were harvested and titered by plaque assay. Virus was never detected in the quadricep samples of animals inoculated sc (Table 4). A single animal inoculated ip (two days post-inoculation) and two mice inoculated iv (at four and six days post-inoculation) had detectable virus in the right quadricep, but the titer was at or just above the limit of detection (6.25 PFU/g tissue). Virus was present sporadically or at low levels in the brain and  
30

serum of animals regardless of the route of inoculation. Virus was detected in the bone marrow of animals regardless of the route of inoculation. However, the presence of virus in bone marrow of animals inoculated sc or ip was more sporadic than animals inoculated iv, where five out of six animals had detectable virus.

5 These results suggest that S55 targets to the bone marrow, especially following iv inoculation.

The level and frequency of virus detected in the serum and muscle suggested that virus detected in the bone marrow was not residual virus contamination from blood or connective tissue remaining in bone marrow samples.

10 The following experiment also suggested that virus in bone marrow was not due to tissue or serum contamination. Mice were inoculated ic with 1200 PFU of S55 in 25  $\mu$ l of diluent. Animals were sacrificed at 0.25, 0.5, 1, 1.5, 2, 3, 4, 5, and 6 days post-inoculation, and the carcasses were decalcified as described in Example 8. Coronal sections taken at approximately 3 mm intervals through the head, spine (including shoulder area), and hips were probed with an S55-specific [ $^{35}$ S]-UTP labeled riboprobe derived from pDS-45. Positive *in situ* hybridization signal was detected by one day post-inoculation in the bone marrow of the skull (data not shown). Weak signal also was present in some of the chondrocytes of the vertebrae, suggesting that S55 was replicating in these cells as well. Although the frequency of positive bone marrow cells was low, the signal was very intense over individual positive cells. This result strongly suggests that S55 replicates *in vivo* in a subset of cells contained in the bone marrow.

15

20

#### EXAMPLE 11

##### Other Sindbis Group Viruses

25 It was of interest to determine if the ability to replicate in the bone marrow of mice was unique to S55 or was a general feature of other viruses, both Sindbis and non-Sindbis viruses, in the Sindbis group. Six 38-day-old female CD-1 mice were inoculated iv with 25  $\mu$ l of diluent containing 10<sup>3</sup> PFU of S55, Ockelbo82, Girdwood S.A., TR339, or TRSB. At 2, 4 and 6 days post-inoculation two mice from each group were sacrificed and whole blood, serum, brain (including brainstem), right quadriceps, and both femurs were harvested for virus titration.

30

The results of this experiment were similar to those with S55. TRSB infected animals had no virus detectable in serum or whole blood in any animal at any time, and with the other viruses tested, no virus was detected in the serum or whole blood of any animal beyond two days post-inoculation (detection limit, 25 PFU/ml). Neither TRSB nor TR339 was detectable in the brains of infected animals at any time post-inoculation. S55, Girdwood S.A., and Ockelbo82 were present in the brains of infected animals sporadically with the titers being at or near the 75 PFU/g level of detection. All the tested viruses were found sporadically at or slightly above the 50 PFU/g detection limit in the right quadricep of infected animals except for a single animal four days post-inoculation with TRSB which had nearly  $10^5$  PFU/g of virus in its quadricep.

The frequency at which the different viruses were detected in bone marrow varied widely, with S55 and Girdwood S.A. being the most frequently isolated (five out of six animals) and Ockelbo82 and TRSB being the least frequently isolated from bone marrow (one out of six animals and two out of six animals, respectively) (Table 4). Girdwood S.A. and S55 gave nearly identical profiles in all tissues. Girdwood S.A., unlike S.A.AR86, is not neurovirulent in adult mice (Example 4), suggesting that the adult neurovirulence phenotype is distinct from the ability of the virus to replicate efficiently in bone marrow.

**TABLE 4**  
**Titers Following IV Inoculation of Virus**

Virus	Animal	Days Post-Inoculation	Tissue Titered			
			Bone Marrow (PFU/ $\mu$ l)	Serum (PFU/ml)	Blood (PFU/ml)	Quadriceps (PFU/g)
SS5	A	2	1125	N.D.*	N.D.	N.D.
	B		488	50	200	N.D.
	A	4	863	N.D.	N.D.	550
	B		113	N.D.	N.D.	
	A	6	N.D.	N.D.	N.D.	50
	B		37.5	N.D.	N.D.	N.D.
	Limit of Detection		37.5	25	25	50
	A		N.D.	N.D.	N.D.	N.D.
TR339	B		1500	75	700	N.D.
	A	4	1050	N.D.	N.D.	N.D.
	B		1762	N.D.	N.D.	400
	A	6	N.D.	N.D.	N.D.	N.D.
	B		N.D.	N.D.	N.D.	N.D.
	Limit of Detection		37.5	25	25	50
	A		N.D.	N.D.	N.D.	N.D.
	B		N.D.	N.D.	N.D.	N.D.
TRSB	A	2	N.D.	N.D.	N.D.	N.D.
	B		N.D.	N.D.	N.D.	N.D.
	A	4	150	N.D.	N.D.	1000
	B		N.D.	N.D.	N.D.	100000
	A	6	N.D.	N.D.	N.D.	N.D.
	B		37.5	N.D.	N.D.	N.D.
	Limit of Detection		37.5	25	25	50

TABLE 4 Continued  
Titers Following IV Inoculation of Virus

Virus	Animal	Days Post-Inoculation	Tissue Titered			
			Bone Marrow (PFU/g)	Serum (PFU/ml)	Blood (PFU/ml)	Brain (PFU/g)
Girdwood S.A.	A	2	22000	2325	1450	30
	B		2500	1200	2600	0
	A	4	788	N.D.	N.D.	N.D.
	B		113	N.D.	N.D.	N.D.
	A	6	N.D.	N.D.	N.D.	N.D.
	B		75	N.D.	N.D.	N.D.
Limit of Detection		37.5	25	25	75	50
Ockelbo82	A	2	N.D.	125	150	N.D.
	B		N.D.	50	500	N.D.
	A	4	N.D.	N.D.	N.D.	200
	B			300	N.D.	N.D.
	A	6	N.D.	N.D.	N.D.	100000
	B		N.D.	N.D.	N.D.	N.D.
Limit of Detection		37.5	25	25	75	50

\* "N.D." indicates that the virus titers were below the limit of detection.

## EXAMPLE 12

Virus Persistence in Bone Marrow

The next step in our investigations was to evaluate the possibility that S.A.AR86 persisted long-term in bone marrow. S51 is a molecularly cloned, attenuated mutant of S55. S51 differs from S55 by a threonine for isoleucine 5 substitution at amino acid residue 538 of nsP1 and is attenuated in adult mice inoculated intracerebrally. Like S55, S51 targeted to and replicated in the bone marrow of 37-day-old female CD-1 mice following ic inoculation. Mice were inoculated ic with 500 PFU of S51 and sacrificed at 4, 8, 16, and 30 days post-inoculation for determination of bone marrow and serum titers. At no time post-inoculation was virus detected in the serum above the 6.25 PFU/ml detection limit. Virus was detectable in the bone marrow samples of both animals sampled at four days post-inoculation and in one animal eight days post-inoculation (Table 5). No virus was detectable by titration on BHK-21 cells in any of the bone marrow samples beyond eight days post-inoculation. These results suggested that the attenuating mutation present in S51, which reduces the neurovirulence of the virus, did not impair acute viral replication in the bone marrow.

It was notable that the plaque size on BHK-21 cells of virus recovered on day 4 post-inoculation was smaller than the size of plaques produced by the inoculum virus, and that plaques produced from virus recovered from the day 8 post-inoculation samples were even smaller and barely visible. This suggests a strong selective pressure in the bone marrow for virus that is much less efficient in forming plaques on BHK-21 cells.

To demonstrate that S51 virus genomes were present in bone marrow cells long after acute infection, four to six-week-old female CD-1 mice were inoculated ic with 500 PFU of S51. Three months post-inoculation two animals were sacrificed, perfused with paraformaldehyde and decalcified as described in Example 8. The heads and hind limbs from these animals were paraffin embedded, sectioned, and probed with a S.A.AR86 specific [<sup>35</sup>S]-UTP labeled riboprobe derived from clone pDS-45. *In situ* hybridization signal was clearly present in discrete cells of the bone and bone marrow of the legs (data not shown). Furthermore, no *in situ* hybridization signal was detected in an adjacent

-47-

control section probed with an influenza virus HA gene specific riboprobe. As the relative sensitivity of *in situ* hybridization is reduced in decalcified tissues (Peter Charles, personal communication), these cells likely contain a relatively high number of viral sequences, even at three months post-inoculation. No *in situ* hybridization signal was observed in mid-sagittal sections of the heads with the S.A.AR86 specific probe, although focal lesions were observed in the brain indicative of the prior acute infection with S51.

TABLE 5

S51 Titers in Bone Marrow Following IC Inoculation of 500 PFU			
Days Post-Inoculation	Titers (Total PFU/Animal)		Limit of Detection
	Animal A	Animal B	
4	2100	380	62.5
8	62.5	N.D. <sup>a</sup>	62.5
16	N.D.	N.D.	62.5
30	N.D.	N.D.	62.5

<sup>a</sup> "N.D." indicates that the virus titers were below the limit of detection.

**Example 13****Replication of S.A.A.R86 within Bone/Joint Tissue of Adult Mice**

Several old world alphaviruses, including Ross River Virus, Chikungunya virus, Okelbo82, and S.A.AR86 are associated with acute and persistent 5 arthritis/arthralgia in humans. Molecular clones of several Sindbis group viruses, including S.A.AR86, were used to investigate alphavirus replication within bone/joint tissue.

Following intravenous inoculation of S.A.AR86 into adult CD-1 mice, viral replication was observed in bone/joint tissue, but not surrounding muscle tissue of 10 the hind limbs. Infectious virus was detectable 24 hrs post-infection; however, viral titer within bone/joint tissue was maximal 72 hours post-infection. Fractionation of hind limbs from infected animals revealed that the hip and knee joints were the predominant sites of viral replication. Replication within bone/joint tissue appears to be a common trait of Sindbis-group viruses, since the laboratory strains TR339 and TRSB 15 also replicated within bone/joint tissue. *In situ* hybridization and S.A.AR86 based double promoter vectors expressing green fluorescent protein were used to further localize S.A.AR86 infected cells within bone/joint tissue. Green fluorescent protein expression was detected in bone/joint tissue for at least one month post-inoculation. These studies demonstrated that cells within the endosteum of synovial joints were the 20 predominant site of S.AAR86 replication.

**WO 98/36779**

**PCT/US98/02945**

**-49-**

**SEQUENCE LISTINGS**

**SUBSTITUTE SHEET (RULE 26)**

-50-

THAT WHICH IS CLAIMED IS:

1. A method of introducing and expressing heterologous RNA in bone marrow cells, comprising:

(a) providing a recombinant alphavirus, said alphavirus containing a heterologous RNA segment, said heterologous RNA segment comprising a promoter operable in said bone marrow cells operatively associated with a heterologous RNA to be expressed in said bone marrow cells; and then

(b) contacting said recombinant alphavirus to said bone marrow cells so that said heterologous RNA segment is introduced and expressed therein.

2. A method according to claim 1, wherein said contacting step is carried out *in vitro*.

3. A method according to claim 1, wherein said contacting step is carried out *in vivo* in a subject in need of such treatment.

4. A method according to claim 1, wherein said heterologous RNA encodes a protein or peptide.

5. A method according to claim 1, wherein said heterologous RNA encodes an immunogenic protein or peptide.

6. A method according to claim 1, wherein said heterologous RNA encodes an antisense oligonucleotide or a ribozyme.

7. A method according to claim 1, wherein said alphavirus is an Old World alphavirus.

8. A method according to claim 1, wherein said alphavirus is selected from the group consisting of SF group and SIN group alphaviruses.

-51-

9. A method according to claim 1, wherein said alphavirus is selected from the group consisting of Semliki Forest virus, Middelburg virus, Chikungunya virus, O'Nyong-Nyong virus, Ross River virus, Barmah Forest virus, Getah virus, Sagiyma virus, Bebaru virus, Mayaro virus, Una virus, Sindbis virus, South African Arbovirus No. 86, Ockelbo virus, Girdwood S.A. virus, Aura virus, Whataroa virus, Babanki virus, and Kyzylagach virus.

5

10. A method according to claim 1, wherein said alphavirus is South African Arbovirus No. 86.

10

11. A method according to claim 1, wherein said alphavirus is Girdwood S.A.

12. A method according to claim 1, wherein said alphavirus is Sindbis strain TR339.

15

13. A helper cell for expressing an infectious, propagation defective, Girdwood S.A. virus particle, comprising, in a Girdwood S.A.-permissive cell:

20

(a) a first helper RNA encoding (i) at least one Girdwood S.A. structural protein, and (ii) not encoding at least one other Girdwood S.A. structural protein; and

25

(b) a second helper RNA separate from said first helper RNA, said second helper RNA (i) not encoding said at least one Girdwood S.A. structural protein encoded by said first helper RNA, and (ii) encoding said at least one other Girdwood S.A. structural protein not encoded by said first helper RNA, and with all of said Girdwood S.A. structural proteins encoded by said first and second helper RNAs assembling together into Girdwood S.A. particles in said cell containing said replicon RNA;

and wherein the Girdwood S.A. packaging segment is deleted from at least said first helper RNA.

-52-

14. The helper cell according to claim 13, further containing a replicon RNA;

said replicon RNA encoding said Girdwood S.A. packaging segment and an inserted heterologous RNA;

5 wherein said Girdwood S.A. packaging segment is deleted from at least one of said helper RNA;

and wherein said replicon RNA, said first helper RNA, and said second helper RNA are all separate molecules from one another.

10 15. The helper cell according to claim 13, further containing a replicon RNA;

said replicon RNA encoding said Girdwood S.A. packaging segment and an inserted heterologous RNA;

wherein said replicon RNA and said first helper RNA are separate molecules;

15 and wherein the molecule containing said replicon RNA further contains RNA encoding said at least one Girdwood S.A. structural protein not encoded by said first helper RNA.

20 16. The helper cell according to claim 13, wherein said first helper RNA encodes both the Girdwood S.A. E1 glycoprotein and the Girdwood S.A. E2 glycoprotein, and wherein said second helper RNA encodes the Girdwood S.A. capsid protein.

17. A method of making infectious, propagation defective, Girdwood S.A. virus particles, comprising:

25 transfecting a Girdwood S.A.-permissive cell according to claim 13 with a propagation defective replicon RNA, said replicon RNA including said Girdwood S.A. packaging segment and an inserted heterologous RNA;

producing said Girdwood S.A. virus particles in said transfected cell; and then

collecting said Girdwood S.A. virus particles from said cell.

-53-

18. Infectious Girdwood S.A. virus particles produced by the method of Claim 17.

5           19. Infectious Girdwood S.A. virus particles containing a replicon RNA encoding a promoter, an inserted heterologous RNA, and wherein RNA encoding at least one Girdwood S.A. structural protein is deleted therefrom so that said Girdwood S.A. virus particle is propagation defective.

20. A pharmaceutical formulation comprising infectious Girdwood S.A. virus particles according to claim 18 or 19 in a pharmaceutically acceptable carrier.

10           21. A helper cell for expressing an infectious, propagation defective, TR339 virus particle, comprising, in a TR339-permissive cell:

(a) a first helper RNA encoding (i) at least one TR339 structural protein, and (ii) not encoding at least one other TR339 structural protein; and

15           (b) a second helper RNA separate from said first helper RNA, said second helper RNA (i) not encoding said at least one TR339 structural protein encoded by said first helper RNA, and (ii) encoding said at least one other TR339 structural protein not encoded by said first helper RNA, and with all of said TR339 structural proteins encoded by said first and second helper RNAs assembling together into TR339 particles in said cell containing said replicon RNA;

20           and wherein the TR339 packaging segment is deleted from at least said first helper RNA.

22. The helper cell according to claim 21, further containing a replicon RNA;

25           said replicon RNA encoding said TR339 packaging segment and an inserted heterologous RNA;

              wherein said TR339 packaging segment is deleted from at least one of said helper RNA;

30           and wherein said replicon RNA, said first helper RNA, and said second helper RNA are all separate molecules from one another.

-54-

23. The helper cell according to claim 21, further containing a replicon RNA;

said replicon RNA encoding said TR339 packaging segment and an inserted heterologous RNA;

5 wherein said replicon RNA and said first helper RNA are separate molecules;

and wherein the molecule containing said replicon RNA further contains RNA encoding said at least one TR339 structural protein not encoded by said first helper RNA.

10 24. The helper cell according to claim 21, wherein said first helper RNA encodes both the TR339 E1 glycoprotein and the TR339 E2 glycoprotein, and wherein said second helper RNA encodes the TR339 capsid protein.

15 25. A method of making infectious, propagation defective, TR339 virus particles, comprising:

transfected a TR339-permissive cell according to claim 21 with a propagation defective replicon RNA, said replicon RNA including said TR339 packaging segment and an inserted heterologous RNA;

producing said TR339 virus particles in said transfected cell; and  
20 then

collecting said TR339 virus particles from said cell.

26. Infectious TR339 virus particles produced by the method of  
Claim 25.

27. Infectious TR339 virus particles containing a replicon RNA  
encoding a promoter, an inserted heterologous RNA, and wherein RNA encoding  
at least one TR339 structural protein is deleted therefrom so that said virus particle  
is propagation defective.

28. A pharmaceutical formulation comprising infectious TR339  
virus particles according to Claim 26 or 27 in a pharmaceutically acceptable carrier.

-55-

29. A recombinant DNA comprising a cDNA coding for an infectious Girdwood S.A. virus RNA transcript and a heterologous promoter positioned upstream from said cDNA and operatively associated therewith.

5           30. An infectious RNA transcript encoded by a cDNA according to claim 29.

10           31. An infectious RNA according to claim 30, said infectious Girdwood S.A. RNA transcript containing a heterologous RNA segment, said heterologous RNA segment comprising a promoter operably associated with a heterologous RNA.

15           32. Infectious viral particles containing an RNA transcript according to claim 30.

20           33. A recombinant DNA comprising a cDNA coding for a Sindbis strain TR339 RNA transcript and a heterologous promoter positioned upstream from said cDNA and operatively associated therewith.

15           34. An infectious RNA transcript encoded by a cDNA according to claim 33.

20           35. An infectious RNA according to claim 34, said infectious Girdwood S.A. RNA transcript containing a heterologous RNA segment, said heterologous RNA segment comprising a promoter operably associated with a heterologous RNA.

36. Infectious viral particles containing an RNA transcript according to claim 34.

## Nucleotide Sequence of S.A.AR86

1 ATTGGGGCG TAGTACACAC TATTAATCA AACAGCCGAC CAATTGCACT ACCATCACAA TCGAGAAGCC AGTAGTTAAC ATAGACGTAG ACCCTCAGAG  
 101 TCCUTTTGTC GTGCAACTGC AAAAGAGCTT CCCGCATTG GAGGTAGTAG CACAGCAGGT CACTCCAAT GACCATGCTA ATGCCAGAGC ATTTTCCAT  
 201 CTGGCCAGTA AACTAATCGA GCTGGAGGTT CCTACACAG CGACGATTT GGACATAGGC ACGGCACCGG CTGGTAGAAT GTTTCCGAG CACCACTTAC  
 301 ATTGGCTTGC CCCCATGCTG AGTCAGAAG ACCGGACCG CATGATGAAA TATGCCAGCA AACTGGCGA AAAAGCATGT AAAGATTACAA ACAAGAACCT  
 401 GCATGAGAAG ATCAAGGACC TCCGGACCGT ACTTGATACA CCGGATGCTG AAACGCCATC ACTCTGCTTC CACAACTGATG TTACCTGCAA CACCGCTGCC  
 501 GAGTACTCGG TCATGCGAGA CGTGTACATC AACGCTCCGG GAACATTTTA CCACCAAGGCT ATGAAAGGGC TCCGGACCT GTACTGGATT CGCTTCCACA  
 601 CCACCCAGT CATGTTCTG GCTATGGCAG GTTGGTACCC TGCATACAAAC ACCAACTGGG CCCACGAAAA AGTCCTTGA GGCGCTAACAA TCGGACTCTG  
 701 CAGCACAAAG CTGAGTGAAG GCAGGACAGG AAAGTGTGG ATAATGAGGA AGAAGGAGTT GGAGCCCGG TGACGGGTTT ATTTCTCCGT TGATGACA  
 801 CTTTACCCAO AACACAGAGC CAGCTTCAG AGCTGGCATE TTCCATCGGT GTTCCACTTG AAAGGAAAGC AGTCGTACAC TTGGCCCTGT GATACTGG  
 901 TGAGCTCGA AGGGTACAGT GTGAGAAAAA TCACCATCG TCCCGGATC ACAGGGAAAAA CGCTGGGATA CGCGGTTACA AACATAGCG AGGGCTTCTT  
 1001 GCTATGCAA GTTACCGATA CAGTAAAAGG AGAACGGGTA CGTGTCCCCG TGTGCGCTA TATCCGGGCC ACCATATGCG ATCAAGATGAC CGCGATAATO  
 1101 GCGACGGATA TCTCACCTGA CGATGCACAA AAACCTCTGG TTGGGCTCAA CCAGGGAATC GTCAATTAAAGC GTAAGACTAA CAGGAACACC AATACCATGC  
 1201 AAAATTACCT TCTCCCAATC ATTGCGACAGG GTTICACCAA ATGGGCAAGG GAGGGCAAGG AAGATCTGA CAAATGAAAAA ATGCTGGGA CCAGAGAGGG  
 1301 CAAGCTTACA TATGGGCTGT GTGTTGGGTTT TGGCACTAAG AAAGTGCACG CTTTCTATCG CCCACCTGGG ACGCAGACCA TCGTAAAGT CCCAGCTCT  
 1401 TTGAGCTGCTT TCCCCATGTC ATCCGTATGG ACTACCTCTG TGCCCATGTC GCTGAGGGAG AAGATGAAT TGCCATTACA ACCAAAGAAG GAGGAAAAAC  
 1501 TGCTGCAAGT CCCGGAGGAA TTAGTTATGG AGGCCAAGGC TGCTTTCAG GATGCTCAGG AGGAATCCAG AGGGGAGAAG CTCCGAGAAG CACTCCACC  
 1601 ATTAGGGCA GACAAGGTA TCGAGGCGC TCGGGAGTT GTCTGGAAAG TGGAGGGCTT CCAGGGGAC ACCGGAGCAG CACTCGTGA AACCCCGCG  
 1701 GGTCTATGAA GGATAATACC TCAAGCAAT GACCGTATGA TCGGACAGTA TATCGTTGTC TGCCGATCT CTGCTGTAAGA GAAAGCTAAA CTGGCACCAG  
 1801 CACACCCGCT ACCAGACCCAG GTTAAAGATCA TAACCGCACT CGGAAGATCA GGAAGGTATG CAGTCGAACC ATACGAGCT AAAGTACTGA TGCCAGCAG  
 1901 AAGTGGCTA CCATGGCCAG AATTCTTAGC ACTGAGTGTAG AGGGCCACCG TTGTTGACAA CGAAAGAGAG TTGTTGAAAC GCAAGCTGTA CCATATTGCC  
 2001 ATGCACGGTC CCCTTAAGAA TACAGAAGAG GAGCAGTACA AGGTTACAAA GGCAGAGCTC CGAGAAACAG AGTACGTGTT TGACGTGGAC AAGAAGCGAT  
 2101 GCGTTAAGAA GGAAGAAGCC TCAAGGACTG TCCCTTCCGG AGAAACTGACC AACCCCGCTT ATCACGAACG AGCTCTTGAG GGACTGAGA CTGGACCCG  
 2201 GTTCCCCTAC AAGGTTGAAA CAATAGGAGT GATAGGCACA CCAGGATCGG GCAAGTCAG TATCATCAAG TCAACTCTCA CGGCACGTGA TCTTGTACCC  
 2301 AGCGGGAAAAG AAAAAAATG CGCGGAATT GAGGGCGAGC TGTACGGGT GAGGGGCATG CAGATCAGT CGAAGACAGT GGATTEGGTTT ATGCTCAACG  
 2401 GATGCCACAA AGCCGTAGAA GTGCTGTATG TTGACGAAGC GTTCCGGTGC CACCGAGGAG CACTACTGC CTGTTGACCA ATCGTCAGAC CCCGTAAGAA  
 2501 GGTAGTACTA TCCGGAGACC CTAAGCAATG CGGATTCTTC AACATGATGC AACTAAAGGT ACATTTCAAC CACCCCTGAA AAGACATATG TACCAAUACA  
 2601 TTCTACAAAGT TTATCTCCCG ACCTGGCACA CAGGGCACTG CCGCTATTGT ATCCGACATG CATTACGATG GAAAATGAA AACACAAAC ECGTGCAGA  
 2701 AGAACATEGA AATCGACATT ACAGGGGCCA CGAACGCCAA CGCAGGGGAC ATCATCTGTG ATGTTTCCG CGGGTGGGTT AAGCAACTGC AAATCGACTA  
 2801 TCCCGACAT GAGGTAAATGA CAGGGGGCGC CTCAACAGGG CTAACCGAAA AAGGAGTATA TCCCTTCCGG CAAAAAGTCA ATGAAAACCC CGTGTACCG  
 2901 ATCACATCG AGCATGTGAA CGTGTGGCTC ACCGGCACTG AGGACAGGGT AGTATGGAAA ACTTTCACGG CGACCCATG GATTAAGCAG CTCACTAACG  
 3001 TACCTAAAGG AAATTTCAAG CCGACCATCG AGGACTGGAA AGCTGAACAC AAGGAAATAA TTGCTGGAT AAACAGTCCC GTCTCCGTA CCAATCCGTT  
 3101 CAGCTGCAAG ACTAACGGTTT GTGGGGGAA AGGACTGGAA CGGATACCTG CGAACGGGGC TATCGTACTT ACCGGTTGCC AGTGGAGCGA GCTGTTCCCA  
 3201 CAGTTTGGGG ATGACAAACCC ACATCCGCC ATCTACCGCT TAGACGTAAT TTGCTTGGCA TTGACCTTGAC AGGGGGCTG TTTTCCAAAC  
 3301 AGAGCATECC GTTAACGTAC CTCCTGGCC ACTCAGGCGA CGCAGTAGCT CATTGGGACA ACAGGCCAGG AACACCCAGG TATGGGTACG ATCACGGCGT  
 3401 TGCGGGCGA CTCTCCCGTA GATTTCCGGT GTTCCAGCTA GCTGGGAAAG GCACACAGCT TGATTTGAG AGGGGAGAA CTAGAGTTAT CTCTGCACAG  
 3501 CATAACTGG TCCCGAGTGAAG CGGCAATCTC CTCACGGCT TAGTCCCCGA CGACAAAGGAG AAACAAACCGG GCGGGTGA AAAATTCCTG ACCAGTTCA  
 3601 AACACCACTC CGTACTTGTG ATCTCAGAGA AAAAATTGAA AGCTCCCCAC AAGAGAATGC ATGGATGCC CGCGATTGGC ATAGCCGGCG CAGATAAGAA  
 3701 CTACAACTG GCTTTCGGGT TTCCGGGCA CGCAGGGTAC GACCTGGTGT TCATCAATAT TGGAACTAAA TACAGAAACC ATCACTTCA ACAGTGGCAA

FIG. 1A

3801 GACCACGGGG CGACCTTGAA AACCCCTTCG CGTTCGGCCC TGAACCTGCC TAACCCCCGA GGCAACCTCG TGGTGAAGTC CTACGGTTAC CCCGACCGCA  
 3901 ATAGTGAGGA CGTAGTCACC CCTCTTGCCA GAAAATTGT CAGAGTGTCT GCAGCGAGGC CAGAGTGCCT CTCAAGCAAT ACAGAAATGT ACCTGATTT  
 4001 CGGACAACTA GACAAACAGCC GCACACGACA ATTCAACCCG CATCATTGA ATTGTGTGAT TTGCTCCGT TACGAGGGTA CAAGAGACGG AATTGGACCC  
 4101 GCACCGTCTG ACCCTACTAA AAGGGAGAAC ATTGCTGATT GTCAAGAGGA AGCAGTTTC AATGCAGCCA ATCCACTGGG CAGACCGGA GAAGGAGCT  
 4201 GCGCTGCCAT CTATAAACGT TGCCGAACA GTTTCACCGA TTCAGCCACA GAGACAGGTAA CGCGAAACCT GACTGTGTGC CAAGGAAAGA AAGTGATCCA  
 4301 CGCGTTGGC CCTGATTTCCG GGAACACCC AGAGGGAGAA GCGCTGAAT TCGTGCAGAA CGCCTACCAT CGAGTGGCG ACTTAGTAA TGACACATAAT  
 4401 ATCAAGTCTG TCGCCATCCC ACTGCTATCT ACAGGCAATT CGCAGGGGG AAAAGACCCG CTTGAGGTTAT CACTTAACGT TTGACAAACC CGCGTAGACA  
 4501 GAAGTGTGCG CGAGCTAACCT ATCTACTGCC TGATAAGAA GTGGAAGGAA AGAATCGACG CGGTGCTCCA ACTTAAGGAG TCTGTAACGT AGCTGAAGGA  
 4601 TGAGGATATO GAGATCGACO ACAGGTTAGT ATGGATCCAT CGCGACAGTT CGCTGAAGGG AAGAAAGGGAA TTCACTGACTA CAAAAGGAAA GTTGTATTG  
 4701 TACTTTGAG GCACCAATT CCATCAAGCA GCAAAGATA TGGCGAGAT AAAGGTECTG TTCCCAAATG ACCAGGAAG CAAACGAACAA CTGTTGCT  
 4801 ACATATTGGG GGAGACCATG GAAGCAATCC CGCAGGGGG AAAAGACCCG CTTGAGGTTAT CACTTAACGT TTGACAAACC CGCGTAGACA  
 4901 TGCCATGAGC CGAAAGGGG TCCACAGACT CAGAAGCAAT AACGTCAAAG AAGTTACAGT ATGCTCTCC ACCCCCCCTTC CAAAGTACAA AATCAAGAAAT  
 5001 GTTCAGAAGG TTCACTGCAC AAAAGTAGTC CTGTTAACCGC CGCATACCC CGCATTGCTT CCGCCCGTA AGTACATAGA AGCACCAAGA CAGCGAG  
 5101 CTCCGGCTGC ACAGGGCGAG GAGGGGGGG GAGTTGAGC GACACCAACA CGACCTGCGAT CTGATAACAC CTGCTTGTAT GTCAAGGACA TCTCACTGG  
 5201 CATGGAAGAC AGTAGGGAAG GCTACTCTT TTGAGGTTT AGCGGATCGG ACAACTACCO AAGGCAAGGTG GTGGTGGCTG AGCTGGATGC CGTCCAAGAG  
 5301 CCTGGGGCTG TTTCACCGCC AAGGCTAAAG AAGATGGCCC GCTGGCGAGC GGCAGAAATG CAGGAAGAGC CAACTCCACCG GGCAGCACC AGCTGTGGG  
 5401 ACGAGTCCCT TCACCTTCTT TTGATGGGG TATCTATATC CTTCGGATCC TTTCAGGACG GAGAGATGGC CGCTTGGCA GGGGACAAC CCCCCGGCAAG  
 5501 TACATGCGCT ACAGGATGTGCT CTATGCTTT CGGATGTTT TCCGACGGAG AGATTGAGGA GTTGGCCCG AGAGTAACCG AGTGGAGGC CGTCTTGT  
 5601 GGGTCATTTG AACCGGGCGA AGTGAACCTA ATTATATCGT CGCGATCAGC CGTATCTTTT CCACCCAGCA AGCAGAGACG TAGACGGCAGG AGCAGGAGGA  
 5701 CGGAATACTG TCTAACCGGG GTAGGGGGT ACATATTTTC GACGGACACA GGGCTGGC ACTTGCAAAA GAAGTCCCTT CTGCAAGAAC AGCTTACAGA  
 5801 ACCGACCTTG GAGGGCAATG TTCTGGAAAG AATCTACGCC CGCGTGTCTG ACACGTCOOA AGAGGAACAG CTCAACTCA GTTACCGAGT GATGCCACCC  
 5901 GAAGCCAACA AAAGCAGGAGT CGACTCTCGA AAAGTAAAGA ACCGAAAGC CATAACCAACT GAGGCACTCC TTTCAGGGT ACTACTGTAT AACCTGCGA  
 6001 CAGATCGCC AGAAATGCTAT AAAGTCACCT ACCGGAAACCC ATCGTATTCC ACCGATGTAC CAGCGAACTA CTCTGACCCA AAGTTTGCTG TAGCTGTTG  
 6101 TAACAACTAT CTGCGATGAGA ATTACCCGAC CGTAGCATCT TATCGATCTA CGCGACGTA CGATGCTTAC TTGGATATGG TAGACGGGAC AGTGGCTTCC  
 6201 CTAGATACGT CAACTTTTG CCCCCCAAG CTAGAAGTT ACCGGAAAG ACACGGATAT AGAGCCCCAA ACATCCCGAG TGGGGTTCCA TCAGCGATGC  
 6301 AGAACACGTT CGAAACAGCTG CTATCTGGCG CGACTAAAGG AAACCTCAAC GTCACACAAA TCCGTGAACG GCAACACTG GACTCAAGCA CATTCACGGT  
 6401 TGATGCTTT CGAAAAATATG CTGCAATGA CGAGTATGG GAGGGTTG CGCGAAAGCC AATTAGGATC ACTACTGAGT TCTGTTACCGC ATACGTGGCC  
 6501 AGACTGAAAG CGCCCTAAGGC CGCCGCACTG TTGCGAAAGA CGCGATAATTG GTTCCCATTG CAAGAAGTGC CTATGGATAG ATTCGTCTAG GACATGAAAA  
 6601 GAGACGTGAA AGTACACCT GGCAAGAAC ACACAGAAGA AAGACGGAAA GTACAAGTGA TACAAGCCG AGAACCCCTG GCGACCGCTT ACCTATGCGG  
 6701 GATCCACCGG GAGTTGATGC CGAGGCTTAC AGCCGGTTTG CTACCCAAACA TTTCACCGCT TTGGACATG TCGGGGGAGG ACTTTGATGC AATCATACCA  
 6801 GAACACTCA AGCAAGGTGA CGCGTGTACTG GAGACGGATA TCCCGTGTGTT CGACAAAGC CAAGACGGC CTATGGCTT ^ACCGGCTG ATGATCTGG  
 6901 AAGACCTGGG TGTGGACCAA CCACACTCG ACTTGATGCA GTGCGGCTTT GGAGAAATAT CATCCACCCA TCTGCCCACG GTTACCGCTT TCAAATTGCG  
 7001 CGCGATGATG AAATCCGAA TGTCTCTAC GCTCTTGTG AACACAGTTC TGAATGTGCT TATGCCACG AGAGTATGG AGGACGGCT TAAAACGTCC  
 7101 AAATGTGCGAG CATTATCGG CGACGACAAC ATTACACG GAGTAGTATC TGACAAAGAA ATGGCTGAGA GTGTTGGCAC CTGGCTAAC ATGGAGGTTA  
 7201 AGATCATGTA CGCAGTCATC CGCGAGAGAC CACCTTACTG CTGGGGTGGAA TTCACTCTGC AAGATTGGT TACCTCCACA CGCTGTCGGC TGGGGACCC  
 7301 CTGAAAGG CTGTTTAAGT TGGGAAACCC GCTCCCGAGCC GCGATGAGC AAGACCAAGA CAGAACAGCC GCTCTGCTAG ATGAAACAAA GCGCTGUTTT  
 7401 AGAGTAGGTA TAACAGACAC CTAGCGATG CGCGTGGCAA CTGGGTATGA GTGAGACAAAC ATCACACCTG TCTGCTGGC ATTGAGAACT TTTGCCAGA  
 7501 CGAAAGAGC ATTCAGGCC ATCAAGGGG AAATAAAGCA TCTCTACGGT GTGCTAAATG AGTCAGCATA GTACATTCA TCTGACTTAAT ACCACAAAC  
 7601 CACCACTATG ATAGAGGAT CTITTAACAT GCTCGGGCCG CGCCCTTCC CAGCCCCAC TCCCATGTGG AGGCGGGGA GAAGGAGGCA GCGGGGGCC  
 7701 ATGCGTCCCC GCAATGGGCT CGCTTCCCAA ATCCAGCAAC TGACCAAGC CGTCACTGCG CTAGTCATTO GACAGGAAC TAGACCTCAA ACCCCACGCC  
 7801 CACCCCCCCC CGCCCTTCCAG AAGAAGCAGG CGCCAAAGCA ACCACCGAG CGGAAGAAAC CAAAACACA CGAGAAGAAG AAGAAGCAAC CGTCAAAAC

Fig. 1B

7901 CAAACCCGGA AAGAGACAGC GTATGGCACT TAAGTTGGAG GCGGACACAC TGTTGGACGT CAAAATGAG GACGGAGATG TCATGGGCA CGCACTGGCC  
 8001 ATGGAAGGAA AGGTAAATGAA ACCACTCCAC GTGAAAGGAA CTATTGACCA CCTTGCTGTA TCAAAGCTA AATTCAACCA ACGTGTGACCA TACGACATGG  
 8101 AGTCGCACA GTTGGCGTC AACATGAGAA GTGAGGCGTT CACCTACACC AGTGAACACC CTGAAGGGTT CTACAACTGG CACCAAGGAG CGGTGCACTA  
 8201 TAGTGGAGGC AGATTTTACCA TCCCCCGGGG AGTAGGGAGC AGAGGAGACA GTGGTGTGTC GATTATGGAT AACTCAGGCC GGGTTGTGCC GATACTGCTC  
 8301 GGAGGGGCTG ATGAGGGAAAC AAGAACCGCC CTTCGGTGG TCACCTGGAA TAGCAAAAGG AAGACAATCA AGACAACCCC GGAAGGGACA GAAGAGTGGT  
 8401 CTGCTGCACT ACTGGTCAGG GCCATGTGCT TGCTTGGAAA CGTGAAGCTTC CCATGCACTT GCGGCCAAC ATGCTACACC CGCAACCAT CGAGACTCT  
 8501 CGACATCTC GAAGAGAACG TGAACACCA GGCCTACGAC ACCTGCTCA ACGCCATATT CGGGTGGGA TCGTGGGCA GAAGTAAAAG AAGCTCACT  
 8601 GACGACTTTA CCTTGACCA CGCGTACTG GGCACATGCT CGTACTGTCA CCATACTGAA CGTGTGTTA GCGCGATTAA GATCGAGCA GTCGGGATG  
 8701 AACGGGACCA CAACACCCATA CGCATACAGA CTTCGGCCA GTTGGGATAC GACCAAAAGG GACGGAGCA CGTAAATAAG TACCCCTACA TGTGGCTGA  
 8801 GCAGGATCAT ACTGTCAAG AAGGCACCATC GGATGACATC AAGATGAGCA CCTCAGGACG GTGAGAAGO CTTAGCTACA AAGGATACTT TCTCTCCG  
 8901 AAGTGTCTC CAGGGGACAG CGTAAACGGT AGCATAGGGA GTAGCAACTC AGCAACGTCA TGCACATGG CCCCAAGAT AAAACCAAAA TCGTGGGAC  
 9001 GGGAAAAATA TGACCTACCT CGCGTTCAGG GTAAAGAGAT TCTTGGACA GTGACCAACG GTGAAAGA AACAACCGCC GGTACATCA CTATGACAG  
 9101 GCGGGGACCG CATGCTATA CATCCTATCT GGAGGAATCA TCAGGGAAA TTACGGGAA CGCACCATEC GGGAGAACAA TTACGTGCA GTGCAAGTGC  
 9201 GGCATTACA AGACCGGAAC CGTACGACG CGTACCGAAA TCACGGCTG CACCGCCATC AAGCAGTGC TCGCTATAAA GACGGACCA ACGAATGGG  
 9301 TCTTCACCTC GCGGACTCG ATCAGACACG CGGACACAC CGCCCAAGGG AAATTCGATT TGCCTTCA CGTGTACCG AGTACCTGCA TGGTCTCTG  
 9401 TCCCCACGG CGGAACGTAG TACACGGCTT TAAACACATC AGCTTCAAT TAGACACAGA CCATCTGACA TTGCTCACCA CGAGGAGACT AGGGGCAAC  
 9501 CGCGAACCAA CCACTGAATG GATCATCGGA AACACGGTTA GAAACCTCAC CGTGCACCGA GATGGCTGG AATACATATG GGGCAATCAC GAACCACTAA  
 9601 GGTCTATGC CCAAGAGTCT GCACCAAGG ACCCTACCG ATGCCACAC GAAATAGTAC AGCATTAATCA TCATGCCAT CGTGTGACCA CCATCTTACG  
 9701 CGTGCATCA CGTGTGTGG CGATGATGAT TGGCGTAACG GTTGCAGCAT TATGTGGCTG TAAACGGCGC CGTGTGTC TGACGCCATA TGCCCTGGCC  
 9801 CCAAATGCCG TGATTCAC TCGCTGGCA CTTTGTGCT GTGTTAGTC GGCTAATGCT GAAACATTCA CGAGGACCAT GAGTTACTTA TGGTGAACA  
 9901 GCGAGCCGTT CTTCGGGTC CAGCTGTGTA TACCTCTGCC CGCTCTCGTC CGTCTAATGC CGTGTGTC ATGCTGCCG CTTTTTAG TGGTGGCCG  
 10001 CGCTTACCTG CGGAAGGTAG ACCGCTACGA ACATGGGACG ACTGTTCCAA ATGTCACCA GATACCGTAT AAGGCACTG TTGAAAGGGC AGGGTACCCC  
 10101 CGCTCAATT TGGAGATTAC TGTATGTC TCGAGGGTT TGCTTCCAC CAACCAAGG TACATTACCT GCAAAATTCAC CACTGTGGTC CGCTCCCTCA  
 10201 AACTCAGATG CTGCGCTCC TTGGATGTC AGGCCGCCG TCACGGACAC TATACTGCA AGGTCTTGG AGGGTGTAC CGCTTCAATG GGGGAGGAGC  
 10301 ACAATTTT TCGGACAGT AGAACACGCA GATGAGTGA CGTACCGTCC AATTTGCACT AGATTCGGCG ACTGACCAAG CGCAGCCAT TAAGGTGCA  
 10401 ACTGCCGCGA TGAAAGTAGG ACTGCGTATA GTGACGGGA ACACCTACCG TTCTCTAGAT GTGTACGTGA CGAGGTCAC ACCAGGAACG TCTAAAGACC  
 10501 TGAAAGTCAT AGCTGGACCA ATTCAGCAT TTGTTACACCC ATTCGATCAC AAGGTCGTTA TCAATGCCGG CGTGTGTC AACTATGACT TTCCGGAAATA  
 10601 CGGAGCGATG AAACCAAGG CGTTGGAGA CATTCAAGCT ACCTCTTGA CTACCAAGA CCTCATGCC AGCACAGACA TTAGGCTACT CAAGCTTCC  
 10701 GCGAAAGAC TGCATGCCC GTACACGGCG GCGGCACTG GATTCGAGAT GTGGAAAAAC AACTCAGGCC GCGCCACTGCA GGAACCGCC CTTTGGT  
 10801 GCGAAGATTGC AGTCAATCGG CTTCGAGCGG TGGACTGCTC ATACGGGAAC ATTCCTCATTT CTATTGACAT CGCAACCGCT CGCTTATCA GGACATCAGA  
 10901 TGACCCACTG GTCTAACAG CGAAATGTA TGTCAGTGA TGCACTTATT CGCGGACTG CGGAGGGATG CGTACCCCTGC AGTATGATC CGACCCGCA  
 11001 GGACAATGCC CTGTACATTC CGATTCGAGC ACACCAACCC TCGAAGAGTC GACAGTTCAT GTCTGGAGA AAGGAGCGGT GACAGTACAC TTACGGACCG  
 11101 CGAGCCCCACA GGCACACTTC ATTGTATGCG TGTTGGTAA GAAGACAAACA TGCAATGCA GATGCAAACCC ACCAGCTGAT CATATCGTA GCAACCCGCA  
 11201 CAAAATGAC CAAGAATTC CGACCGGCT ATCAGGAAAC TCTAATGCA CGTGTGTTGC CGCTTCCGCG CGCTTAAATAT TATAGGACTT  
 11301 ATGATTTTG CTTCGACCAT GATGCTGACT AGCACCGAA GATGACCGCT AGCCCGAAC GACGGGACCA GCGAAACCTG AGTACTTCC GAGGAACG  
 11401 TGTCATAAT GCACTGGCT GTTATTAATG ATCCCGCTT ACCGGGGCA ATATAGCAAC ACCAAACACTC GACGGTATTC CGAGGAAGGG CAGTGCATAA  
 11501 TGCTGGCGAG TGTGGCAAAT TAATCACTAT ATTAACCAATT TATTCAGGGG AGGCCAAAC TCAATGTATT TCTGAGGAAG CATGGTCAT AATGCCATGC  
 11601 ACCGTCGCA TAACCTTTA TTATTTCTT TATTAATCA CAAAATTTG TTTTAAACAT TTC

Fig. 1c

## S.A.AR86

**A.** Amino Acid Sequence of the Nonstructural Polyprotein

1 MEKPVNVYDV DPOSPFVYQL QKSFPQPEVV AQCVTPNDHA NARAFSLAS KLIIELEVPTT ATILDIGSAP 4RMPSEHQY HCVCPCMRSFB DPDRMMKTYAS  
 101 KLAEKACKIT NKNLHEKKD LRTVLDITPA STPSLCPHND VTCNTRAEYS VMQDVYINAP GTIYHQAMKG VRTLYWIGDF TTQFMPSAMA QSYFAYNTNW  
 201 ADEKVLEARN IGLCSTKLSE GRGTGELSIR KKEELPGSRV EHRASLQSWH LPSVYFLHKKG GSYTCCRCDTV VSCEOYVYICK ITSPGNTGE  
 301 TVGYAVTNNS EGFLLCKVTD TVKGERVSFI VCTYPATIF DQMTGIMATO ISPDQAQKLL VGLNQQRIVIN GKTNRNTNM QNYLLPMQ QPSKWAERK  
 401 EDLDNEBKMLG TREKRLTYGC LWAFRTKKVH SFYRPGQTQ IVKVPASFA FPMSSIVWTTL PMMSLRQKMK LALQPKKEEK LLQVZEELVM EAKAALPDAQ  
 501 EESRAEKLRS ALPPVADKQ IEAAAEVQVE VEGLQADFTQ ALVETPYRGHV RNPQANDRM RGPQVSPF SVLKNAKLF AHPLADQVKI ITISGRSGRY  
 601 AVEPYDALKV MPAGSAVFW EFLALSATV LVNVYREFVN RKLHYHAMING PAKNTTEEFVY KVTKAELAEST EYVFDVDKKA CYCCKERASGL VL5GELTHPP  
 701 YHELALEGLK TRPAVPMYVE TIGVIGCTPS GKSAAKSTV TARDLVTSGK KENCREEAD VLRLRGMQIT SKTVDSVMLN OCHKAVEVLY VDEAFRCHAG  
 801 ALLLAJIAVR FRKVVVLGGD PKQCGFFNNMM QLKVFHNPK KDCITKTFYK FISRCTQPV TAIVSTLHYD GKMKTTNPCK KNEDIDTGK TKPKPMQDIL  
 901 TCFROWWVEQL QIDYDGHEMV TAAASQGLT KGVYAVRQKV NENPLYAITS EHVNVLLTIT EDRLVWVKTQ GDWPWIKQLTN VPKGPNQATI EDWBAEHHGI  
 1001 IAANSPAPR TNPSFCKTMY CWAKALEPIL ATAGIVLTGQ QWSELPFQPA DDKPHSAIYA LDVICDQFPO MDLTSGLFSK QSIPLTYHPA DSAAPVAHWD  
 1101 NSPOTRKGY DHAYAAELSR RPPVFLQLAGK QTQDLQGTA TRVISAOHNL VPVNRLNPLHA LVEHKEKOP GPVKEKPLSFQ KHRSVLVISE XIEAPHKRI  
 1201 EWIAPIGAG ADKYNMNLAFG FPPQARYDLY FPIGTYKRN HMFQOCEDHA ATLXTLSRSA LNCLNPGTL VVKSYGYADR NSEDVUTALA RPKVYVSAAR  
 1301 PECVSYNTEM YLIPRQLDNV ATROFTPHHL NCVISSYTED TRDGVGAAPF YRTKRENIAD COEBAVVNAA NPLGRPGEGV CRAIKRWPW SPIDSATBTO  
 1401 TAKLTVCQG KVHVAVGDPF RKHPEAEALKL LLQNAHVAVA DLVNEHNIKS VAIPLSTLTA YAAGKDRLEV SINCLETAL RTDADVITYC LDICKWKEERD  
 1501 AVLQLKESVT ELKDEDMEID DELWYHIPS CLKGRKGKFST TKGKLYTYPE GTKPHQAAKD MASEKVLFWN DQESNEQLCA YILGETVIMEAI REKCPVDHNP  
 1601 SSSPPTKFLC LCAMYAMTPC VHLRLRSNNVK ETVCSSTPL PTKYKIKONVK VQCTKVVLFN PHTPAPVPAR KYIEAPZQPA APPAQAEZAP GVVATTPPA  
 1701 ADNTSLDVDT ISLDMEDSSE GSFLSSFGS DNYRQXQVVA DVHVAEQEPF VFFFPLKKA BAAARMQEE PTTFASTSSA DESHLHSFDQ VSI5FOSLFD  
 1801 GBMARLAAAG PPASTCTPDV PMSCGFSRSDO ETPFJSRRT ESEPVLGSPF EPGEVNSIS SRAVSPPRR QRERRRSRR TEYCLTGVGQ YIFSTDITGQG  
 1901 HLOQQKSYLVQN QLTETPLERN VLERIYAPL DTKEEQLKL RYQNMPIETAN KSYRQSKRVE NQKAETTERI LSGILRLYNSA TDOPECYKIT YPKPBYSSV  
 2001 PANYSDPKPA VAVCNLYHE NYTYPASYQI TDEDAYLDM VDGTVACLDT ATFCPAKLAS YPKRHEYRAP NIRSAPVPSM QNTLQNVJIA ATKRNCHVYQ  
 2101 MRELPTLDSA TNVECPKRY ACDETYWEEP ARKMRITTE FTVAYVARLK GPKAALFAK THNLVPLQEV PMDRFVMDMM RDVKVTPGTR HTEERPKVQV  
 2201 QAAAEPLATA YLCGIBRRELV RRLTAVLPN IHTLDMASB DFDAIAEHEF KQGDPVLETD IASFDKESODD AMALTGLMIL EDLGVDQPLL DLICAGEI  
 2301 STHLPTGTR FKFOAMMKSG MFLTLPVNTV LNVVIAESVY EBRLLTKSKCA AFICDDMIH GVYSDEKEMAS RCATWLMMEV KIDAVIGER PPYPCOGF  
 2401 QDSVITACK VADPLKLRLPK LGKPLPADDE QDEDRRALL DETKAWFWRG ITDTLAVAYA TRYEVNDITV VLLALRTPAQ SKRAFPQAIRG EKHLYGGPK

**B.** Amino Acid Sequence of the Structural Polyprotein

1 MNRGFFNNMQL XRPFPPTAM WRPRRRRQAA PMPARNGLAS QQQLTTAVS ALVIGQATR QTPRPRPPR QKKQAKPQPP KPKKPKTQEK KKKQPAKPKP  
 101 GKKQRMALKL EADRLFDVKN EDGDVIGHAL AMEGKVMKPL HVKGTDHPV LSKLKFTKSS AYDMEFAQLP VNMRSEAFTY TSEHPEGFYN WHHOAIVQYSG  
 201 CRPTPREGV GRGDGSRPVM DNGCRVVAIV LGGADEGTTRT ALSVTWNSK GKTKEITPTEG TEEWSAALPV TAMCLLGNVS FPCNHRPTCY TREPSSLADI  
 301 LEENVNHEAY DTLLNAILRC GSSEGRSKRSV TDDFTLTSPI LGTCYCHRT EPCPSKPIKE QVWDEADONT IRIOTSAQFG YDQSGAASN KYRTYMSLEQD  
 401 HTVKEGTMDD IKISTSGPCR RLSYKGYFLL AKCPGDSVT VSIASSNSAT SCTMARKXP KFVGREKYDL PPVHGKKKPC TVYDRLKEETT AGYITMHKPO  
 501 PHAYTSTYLER SSCKVYAKPQ SGKNTTYECK CDDYKTTOTVT TATEITGTA IKGCVATKSD QTKWVFRSPD SIRHADHTAQ GKLHLPKLI PFTCMVPAH  
 601 APNIVVHGFKH ISLQLDTDHL TLTLTTRRLGA NPETTEWII GNTVNRPTVD RDLGEYIWGN HEVRYVVAQE SAPGDPHGWV HEIVQHMYHKA HVVTTILAVA  
 701 SAAVAMMIGV TVAALCAKKA RRECLITPAL APNAVITPSL ALLCCVRSAN AEITETEMSY LWNSNQSPFFV VOLCIPLAAV VVLMRCCSSC LPFLVVAGAY  
 801 LAKVDAYEKA TTVPVNPQIP YKALVERAGY APLNLEITVM SIEVLSTNQ EYITCKPTTV VSPPKVRCGG SLECOAPAHA DYTCCKVFGGV YPFMWGGAQC  
 901 FCDSSENSQMS EAYVELSVDC ATDHAQAKV HTAAKVGRL RVYGRNTSFL DVYVNGVTPG TSKDLKVIAG PISALFTTFD HXVVIN&GLV YNYDPPEYGA  
 1001 MKPOAEGDIQ ATSLTSKDLI ASTDIRLRLKP SAKNVHVPPY QAASGFEMWK NSSGRPLQET APFGCKIAVN PLRAVDCSYG NIPSIDPNA AAFIKTSDAF  
 1101 LYSTYKCDVS ECTYSAFDGG MATLQYVSDR EGOCPVHSN STATLQESTV HYLEKGAFTV HESTASQAN FIVSLCGKKT TCNAECKPA DIVSTPHK  
 1201 DQEPQAAISK TSWSWLFALF GGASSLILG LMIFACSMML TSTRR

FIG. 2

## Nucleotide Sequence of Girdwood S.A.

1 NTTONCGGG TAGTATAACAC TATTGAATCA AACAGCCGAC CAATTGCACT ACCATCACAA TGGAGAAGCC AGTATTTAAC GTAGACGTAG ACCCCGAGAG  
 101 TCCCTTTGTC GTCCAACCTGC AAAAGAGCTT CGCGCAATTG GAGGTAGTAG CACAGCAGGT CACTCCAAT GACCATGCTA ATGCCAGAGC ATTTTCCAT  
 201 CTGGCCAGTA AACTAATCGA GCTGGAGGTT CCTACCACAG CGACGATTTT GGACATAGGC AGGGCACCGG CTGGAGAAT GTTTGGAG CACCAAGTAC  
 301 ATTCCGTTTCCCCATCGGT AGTCCAGAAG ACCEGGACCO CATGATGAAA TATGCCAGCA AACTGGCGA AAAAGCATGC AAAGATTACGA ATAAGAACCT  
 401 GCATGAGAAG ATCAAGGACC TCGGGACCGT ACTTGATACA CGCGATGCTG AAACGGCCATC ACTCTGCTTC CACAAACGATG TTACCTGCAA CACCGGTGCC  
 501 GAGTACTCCG TCATGCAGGA CGTGATACATC AACGCTCCCG GAACATTTA CCATCAGGCT ATGAAAGGGG TGCGGACCT GTACTGGATT GGCTTCGATA  
 601 CCACCCAGTT CATGTTCTCG 5CTATGGAG GTTCGTACCC TGCGTACAC ACCAACTGCG CGACGAAAAA AGTCTCGAA CGCGTAAACA TCGGACTCTG  
 701 CACCAACAACTGAGTGAAG CGACGAGAG AAAGTGTGCGA ATATGAGGA AGAAGGAGTT GAAGGGCGG TGACGGGTTT ATTTCTCCGT TGGATGCGACA  
 801 CTTTACCCAG AACACAGAGC CAGCTTGCG AGCTGGCATC TTCCATCGGT GTTCCACCTG AAAGGAAAGC AGTCGTACAC TTGCGGCTGT GATACAGTGG  
 901 TGAGCTCGA AGGCTACGTA GTGAGAGAAA TCACCATCG TCCGGGATC ACAGGAGAAA CGTGGGATA CGCGGTTACA AACATAGCG AGGGCTCTT  
 1001 GCTATGCAAA GTTACCCGATA CAGTAAAAGG AGAACGGGTA TCGTTCCCCG TGTGACGUTA TATCCGGCC ACCATATGCG ATCAGATGAC CGGCATAATO  
 1101 GCGACGGATA TCTCACCTG CGATGCACAA AAACCTCTGG TTGGGCTCAA CGACGGAATC GTTCAATTACG GTAAGACTAA CAGGAACACC AATACCATGC  
 1201 AAAATTACCT TCTGCCAATC ATFGCACAAAG GGTTCGAAATGGGCAAG GAGCGGAAAG AAGACCTTGA CAAAGAAAAA ATGCTGGGTA CGAGAGAGCG  
 1301 CAAGCTTACA TATGGCTGT TGCGCACTAAG AAAGTGCACG CGTTCATCGG CCCACCTGGG ACAGCAGACCA TGCTAAAAGT CCCAGGCTCT  
 1401 TTAGGGCTT TCCCCATGTC ATCCGATGG ACTACCTCTT TGCCCATGTC CGTGAGGAG AGATAAAAAT TGCGATTACA ACCAAAGAAG GAGGAAAAC  
 1501 TGCTGCAAGT CCGGGAGGAA TTAGTCATGG AGGCAAGGC TGCTTTCGAG GATGCTCAGG AGGAATCCAG AGGGAGAAG CTCCGAGAAG CACTCCACC  
 1601 ATTAGTGCGCA GACAAAGGT TCGAGGAGC CGCGGAAGTT GTCTGGAG TGAGGGGCT CGAGGGGGAC ATCGGAGCAG CACTGTCGA AACCCCGCCC  
 1701 GGTCACTGAA CGATAATACC ACAAGCAAAAT GACCGTATGA TCGGACAGTA CATGTTGTC TCGGCAACCT GTGCTGAA GAACGGTAAA CTGGCACCAAG  
 1801 CACACCCGCT AGCAGACAG GTTAAGATCA TAACGCACTC CGGAAGATCA GGAAGGTATG CAGTCGAACC ATACGAGCT AAAGTACTGA TGCCAGCAG  
 1901 AAGTGGCGTA CCATGGCCGAA AATTCCTAACG ACTGAGTGAAG AGGGCCACCC TAGTGTACAA CGAAAGAGAG TTGTGAAACC GCAAGCTGTA CCATATTGCC  
 2001 ATGACCGGTC CGCGTAAGAA TACAGAAGAG GAGGAGTACA AGGTTACAAA GGCAGAGCTC CGAGAAAGAG AGTACGTGTT TGACGGGAC AAGAAGCGAT  
 2101 GCGTCAGAA GGAAGAAGCC TCAGGACTTGT TCTCTCGGG AGAACTGACCC AACCCCGCTT ATCACGAACT AGCTCTTGAAG GGAAGTGAAGA CTGGACCGT  
 2201 GGTCCCGTAC AAGGTTGAAA CAATAGGAGT GATAGGGGCA CGAGGATCGG GCAAGTGGCC TATCATCAAG TCAACTGCA CGGCACGTGA TCTTGTGTTAC  
 2301 AGGGAAAGA AAGAAAATG CGCGGAAATT CAGGGCGATGG TGCTACCGCT GAGGGGCGATG CAGATCACGT CGAAGACAGT GGATTEGGTTT ATGCTCAACG  
 2401 GATGCCGCAA AGGGTAGAA GTGCTGTATG TTGACGGAGC GTTGGGGTCC CACGGAGGAG CACTACTTGC CTGATTGCA ATCGTCAGAC CGGGTCTATAA  
 2501 GGTAGTGTCA TGCGGAGACCT TAAGCAATG CGGATCTTC AACATGATGC AACTAAAGGT ATTTCTAAC CACCCGAAA AAGACATATG TACCAAGACA  
 2601 TTCTACAAGT TTATCTCCG ACCTGGCACA CGGCCAGTCA CGGCTATTGT ATCGCACACTG CATTACGATG GAAAATGAA AACACAAAC CGCTGCAAGA  
 2701 AGAACATCGA AATCGACATT ACAGGGGCCA CGAAGCGAA CGCAGGGGAC ATCATCTGA CATGCTTCCG CGGGTGGGTT AAGCAACTGC AAATCGACTA  
 2801 TCCCGGACAT GAGGTAAATGA CGGGGGGGCCTCACAAGGGAA TAACCGAGAA AAGGAGTATA TCCCGTCCGG CAAAGGTCA ATGAAAACCG GCTGTACCGG  
 2901 ATCACATCG AGCATGTGAA CGTGTGCTC ACCCGCACTG AGGACAGGGT AGTATGGAAA ACTTTACAGG CGCACCATG GATTAACCGAG CTCACTAAGC  
 3001 TACCAAAAGG AAATTTCAAA GCCACCATCG AGGACTGGGA AGCTGAAACAC AAGGGATAAA TTGCTGCGAT AACAGTCCC GCTCCCGTA CCAATCCGTT  
 3101 CAGCTGCAAG ACTAACCTTGT GCTGGGCAAG CGGACTGGG CGGATACTGG CGACGGGGGG TATGCTACTT ACCGGTTGCC AGTGGAGCGA GCTGTCCCA  
 3201 CAGTTTCAAG ATGACAAACCG ACACCTGGCC ATCTACGCCG TGAGCGTAAAT CTGCTTAAAG TTTCGGCA TGCGCTTGAC AAGGGACTG TTTCCAAAC  
 3301 AGAGCATECCG GTTAACTGTCAC TCTCTGGCG ATTCAAGCGG CGCAGTAGCT CATTGGACAA ACAGGGCAGG AACCCGCAAG TATGGGTACG ATCACGGCGT  
 3401 TGCCGGCAA CTCTCCGTA GATTTCCGGT GTTCCAGTCA CGTGGGAAAG GCACACAGCT TGATTTGGAG CGGGCAGAA CTAGAGTTAT CTGGCACAG  
 3501 CATAACCTGG TCCCAGTGAAG CGCGAACCTTC CGCGCACGGCT TAGTCCCGGA CGACAGGGAG AAGCAACCCG GCGGGTCAA AAAATTCTTG AGCCAGTTC  
 3601 AACACCAACTC CGTACTGTG TGCTCAGAGG AAAAATGAA AGCTCCCCAC AAGAGAATCG AATGGATGCC CGCGTGGGC ATACCCGGG CTGATAAGAA  
 3701 CTACAACTG GCTTCCGGT TTCCGGCGA CGCACCGTAC GACCTGGGT TTATCAATAT TGGAACATAA TACAGAAACC ATCACTTCA CGAGTGGGAA

Fig. 3A

3801 GACCATGCCG CGACCTTGAA AACCCCTCTG CTTTCGGCCC TGAACCTGCCT TAACCCCGA GGCACCCCTG TGGTAAAGTC CTACGGTTAC GCGAACCGCA  
 3901 ATAGTGAGGA CTTAGTCACC GCTCTTGCCT GAAAATTGTG CAGAGTGTCT GCAGCGAGGC CAGAGTGCCT CTCAGGAAT ACAGAAATGT ACCTGATCTT  
 4001 CGCAGAACTA GACAACAGCC GCACACGACA ATTCAACCGG CATCATCTGA ATTGTGAT TCTCTCCCTA TAGGAGGATA CAAGAGACGG AGTTGGAGCC  
 4101 GCACCGTCT ACCECCACTAA AAGGGAGAAC ATTTGCTGATT GTCAAGAGGA AGCAGTTGTC AATGCCGCA ATCCGCTGGG CAGACCCAGG GAAAGGATCT  
 4201 GCCGTGCCAT CTATAAACGT TGGCCGAACA GTTTCACCGA TTCAAGGCCA GAGACCCGGCA CGCAGAAACT GACTGTGTC CAAGGAAGA AAGTGATCCA  
 4301 CGCGGTTGGC CCTGATTCTT GGAAACACCC AGAGGCAGAA GCGCTGAAAT TCTCTGAAAA CGCTTACCAT GCAGTGCAG ACTTTGTAAGA TGAACTATAAT  
 4401 ATCAAGTCTG TCGCCATCCT ACCTGCTATCT ACAGGCATTG ACAGCACGGG AAAAGACCCG CTTGAGTAT CACTTAACTG CTGACAAACC CGCTAGATA  
 4501 GAACTGATGC GGACGTAACC ATCTACTGCC TGGATAAGAA GTGGAAGGAA AGAATCGAGG CGGTGCTCA ACTTAAGGAG TCTGTAATAAG AGCTGAAAGA  
 4601 TGAGGATATG GAGATGCCG AGCAGTTAGT ATGGATCCAT CGGAGACAGT GCTCTGAGGG AAGAAGGGG TTCACTACTA CAAAAGGGAA GTTGTATTCC  
 4701 TACTTTGAGG CGACAAATT CCATCAAGG CGAAAGATA TGGCGAGAT AAAGGCTCTG TTCCCAATG ACCAGGAAAG CAAGGACCAA CTGTGTGCT  
 4801 ACATATTGGG GGAGACCATG GAAGCAATCC CGGAAAGATG CGCGGTCGAC CACAACCCGT CGCTAGGCC GCGAAAAACG CTGCCGTGCC TCTGCTATGTA  
 4901 TCCCATGCC CGAGAAAGG TCCACAGACT CAGAAGCAAC AACGTCAAAG AAGTTACAGT ATGGCTCTCC ACCCCCCCTTC CAAAGTACAA ATCAAGAAC  
 5001 GTTCAAGAAGG TTCAGTGCAC AAAAGTAGTC CTUTTTAACG CGCATACCCG TGCTTCTGTT CGCCCGGCGA AGTACATAGA AGGCCAGAGG CAGCCGCG  
 5101 CTCCCGCTGC ACAGGGCGAG GAGGCCCCCG AAGTTGAGC AACACCAACA CCACCTGCAG CTGATAACAC CTCCCTGAT GTCAAGGACA TCTCACTGAA  
 5201 CATGGAAGAC AGTAGCGAAG GTCACTCTT TTGAGCTTT AGCGGATCGG ACAACTCTAT TACTAGTATG GACAGTTGGT CGTCAGGACC TAGTTCACTA  
 5301 GAGATAGTAG ACCGGAAGGCA GGTGGTGGTG GCTGACGTCG ATGCCGTCGA AGAGCCCTGCC CCTGTTCCAC CGCCAAAGCT AAAAGAAGATG GCGCGCTGG  
 5401 CAGCGGCAAG ATGCGAGGA GAGCAACTC CACCGGGCAAG CACCGCTCT CGGGACGAGT CCTTCACCT TTCTTTGGT GGGGTATCCA TGTCTTCCG  
 5501 ATCCCCCTTC GACCGGAGAGA TGGCCGCTT GGCGCGCGA CAACCCCCCGG CAAAGTACATG CCTCTACGGAT GTCCCTATGT TTTCGGATC GTTTCCGAC  
 5601 GGAGAGATTG AGGAGCTGA CGCGAGAGTA ACCGAGTCTG AGCCCCCTCTT TTGGGGTCA TTGAAACCGG CGGAAGTGA CTCAATTATA TGTCGGAT  
 5701 CAGTTGTATC TTTCACCA CGCAAGCAGA GACGTAGACG CAGGAGCAGG AGGAGGAAAT ACTGACTAAC CGGGGTAGGT GGGTACATAT TTGCGACCGA  
 5801 CACAGGCCCT GGGCACTTC AAATGGAGTC CGTTCTGCAG AATCAGCTTA CAGAAGGAC CCTGGAGCGC AATGTTCTGG AAGAATCTA CGCCCCGGG  
 5901 CTCGACACGT CGAAAGAGGA ACAGCTCAA CTCAGGTACG AGATGATGCC CACCGAAGC AACAAAGGA GTTACCGTC TAGAAAAGTA GAAAATCAGA  
 6001 AAGCCATAAC CACTGAGCGA CTGCTTCACT GGCTACGACT GTATAACTCT CCCACAGATC AGCCAGAAATG CTATAAGATC ACCTACCCG AACCACCTGA  
 6101 TTCCACCGATG GTACCGGGCA ACTACTCTG CCGAAAGTTT GTCTGATGCTG TTGCAACAA CTATCTCCAT GAGAAATTACG CGACGGTAGC ATCTTATCAG  
 6201 ATCACCGACO AGTACGATGC TTACTTGGAT ATGGTAGACG GGACAGTGCCT TTGCTTAGAT ACTGCAACTT TTGCCCCCGE CAAGCTTAAAG AGTACCCCGA  
 6301 AAAGACACGA GTATAGAGCC CCAAACACTC CGAGTGCCTG TCCATCAGCG ATGCAGAACG CGTTGAAAGA CGGTGCTATT GCGCGACTA AAAGAAACTG  
 6401 CAACCTACA CAAATGGCTG AATTGCAAC ACTGGACTCA CGGACATTCA CGTTGAATG CTTCGAAAA TATGCATGTA ATGACGAGTA TTGGGAGGAG  
 6501 TTGCCCCGAA AGCCAATTAG GATCACTACT GAGTTCTGTA CGGCATACGT GCGCAGACTG AAAGCCCCTA AGGCCGCGCG ACTGTTGCAG AAGACGCTA  
 6601 ATTTGGTCCC ATTCCAAGAA GTGCCCTATGG ATAGGTTCTG CATGGACATG AAAAGAGACG TGAAAGTTAC ACCTGGCAGG AAACACACAO AAGAAAGAC  
 6701 GAAAGTACAA GTGCTACAG CGCGAGAACCC CCTGGCGACC GTTACCTGT GCGGGATCCA CGGGGAGTTA GTGCGCAGGC TTACAGCCGT TTGCTACCC  
 6801 AACATTACA CGCTTTTGATGAGCTGGG CAGGACTTTG ATGCAATCAT AGCAGAACAC TTGAGCAAG GTGACCCGGT ACTGGAGACG GATATCGCT  
 6901 CGTTGACAA AGCCAAGAC GACGCTATGG CTAACTGCG CCTGATGATC TTGGAAGACC TGGTGTGGA CCAACCACTA CTGACTTGA TCGAGTGC  
 7001 CTTTGGAGAA ATATCATCCA CCCATCTGCC CACGGGTACG CGTTCAATG TCGGGGGAT GATGAAATCC GGAATGTTCC TCACGGCTTT TTGCAACACA  
 7101 GTTCTGAATG TCGTTATGCC CACCGAGAGTA TTGGAGGAGC CGCTTAAACG GTCTAAATGT CGACGATTTA TCGGGGACGA CAACATCATA CACGGAGTAG  
 7201 TATCTGACAA AGAAATGCTG GAGAGGTGTG CCACCTGCTG CAAACATGAG GTTAAAGATCA TTGACCCAGT CACGGGGAG AGACCCCCCTT ACTTCTUCC  
 7301 TGGATTCTAC TTGCAAGATT CGGTTACCTC CACAGCGTGT CGCGTGGCGG ACCCCCCGAA AAGGCTGTTT AAGTTGGGT AACCGCTCCC AGCCGACGAC  
 7401 GACCAAGACG AAGACAGAGC ACCGGCTCTG CTAGATGAA CAAAGGGCTG GTTACAGATA GTTACGAGA ACACCTTACG AGTGGCCGCG GCAACTCGCT  
 7501 ATGAGGTTAGA CAACATCACA CCTGCTCTGC TGGCATGAG AACTTTGCG CAGAGCAAA GACGATTTCA AGCCATCAGA GGGGAAATAA AGCATCTTA  
 7601 CGGTGGTCTT AAATAGTCAG CATAGCACAT TTGATCTGCAT TAATACCAAC ACACCCACAC CATGAATAGA GGATCTTAA ACATGCTGG CGCCCGCCCC  
 7701 TTCCCGCCCC CCACTGCCAT GTGGAGGCCG CGGAGAAGGA GCGAGGGCC CGCGATGCCG CTCCGCAATG GGCTGGCTTC CCAAACTCCAG CAACGACCA  
 7801 CACCCGTCAG TGCCCTGATC ATGGACAGG CAACTAGACG TCAAAACCCCA CGCCCGACGCC CGGGCGCGCG CCAGAAGAAG CAGGGCGCAA AGCAACCCAC

FIG. 3 B

701 GAAGCCGAAG AAACCAAAAA CACAGGAGAA GAAGAAGAAG CAACCTGC AAACCAAACC CGGAAGAGA CAACGTATGG CACTCAAGTT GGAGGCCAAC  
 8001 AGACTGTTCG ACGTCAAAAA TGAGGACGGA GATGTCATCG GGACGCCT GCCTCATGGAA GGAAAGGTAA TGAACCAACT CCACGTGAAA GGAACATATTG  
 8101 ACCACCCCTGT GCTATCAAGG CTCAAAATCA CCAAGTCCTC ACCATAACAC ATGGAGTTTG CACAGTTGCC GGTCAACATG AGAACTGAGG CCTTCACCTA  
 8201 CACCAAGCGA CACCCCTGAAG GTTTTACAA CTGGCCACAC GGAGCGGTGC AGTATAGTG AGGTAGATT ACCATCCCCC CGGGAGTAGG AGGGAGAGGA  
 8301 GACAGTGTTC GTCGGATTAT GGATAACTCA GGCGCGGTTTG TGCGGATAGT CCTCGGGAGGG GGTGATGAGG GAACAAGAAC TCCCCTTTCG GTCGTACCT  
 8401 GGAATAGCAA AGGAAAGACA ATCAAGACAA CCCCGGAAGU GACAGAAGAG TGUTCTCGAG CACCACTGGT CACGGCCATG TGCTTGTCTG GAAACGTGAG  
 8501 CTTCCCATGTC AATCGCCCGC CCACATGCTA CACCEGGCGA CCATCCAGAG CCTCTTGACAT CCTGGAAGAG AACGTGAAACCC ACAGGGCTTA CGACACCGT  
 8601 CTCAACCCCA TATTGGGTG CGGATCGTCC CGGAGAAGCA AAAGAAGCGT CACTGACGAC TTACCTTGA CCACCCGTA CTGGGCACAA TGCTCGTACT  
 8701 GTCACCATAC TGAACCGTGC TTTAGCCCCA TTAAGATGCA GCAGGTCTGG GATGAAGCGG AGCACAACAC CATAACGCTA CAGACCTCCG CCCAGTTTG  
 8801 ATACGACCAA AGCGGAGCGAG CAAAGCTCAA TAAGTACCCG TACATGTGCG TCGAGGAGGA TCATACCTC AAAGAAGGCA CTATGGATGA CATCAAGATC  
 8901 AGCACCTCAG GACCGTGTG AAGGCTTACG TACAAAGGAT ACTTTCTCC CGCGAAGTGT CCTCCAGGGG ACACCGTAAAC GGTAGTATA CGGAGTAGCA  
 9001 ACTCAGCAAC GTCATGCACA ATGGCCCGCA AGATAAAACCA AAAATTCTGG GGACGGGAAA AAATATGACCT ACCTCCCGT CACGGTAAGA AGATTCTTG  
 9101 CACAGTGTAC GACCGTGTGA AAGAAAACAC CGCGCGCTAC ATCACTATGC ACAGGGCGG ACCGACGCC TATACTGCT ATCTGGAGGA ATCATCAGGG  
 9201 AAAGTCTAG CGAAGGCCAC ATCGGAAAG AACATTACGT ACAGGTGCAA TGCGGGCGAT TACAAGACCG GTACCGTTAC GACCCGTACC GAAATCACGG  
 9301 GCTGCACCC CATCAAGCG TGCGTGTGCT ATAAGAGCGA CCAAACCGAAG TGGGTCTTCA ATTGCGGGA CTTGATCAGA CATGCCAAC ACACGGCCA  
 9401 AGGGAAATTG CATTACCTT TCAAGGTGTAT CGCGAGTACG TGCGATGGTCC CTGGTGGCCA CGCGCGGAAC GTAGTACACG CCTTTAAACA CATCGCTC  
 9501 CAATTAGACA CAGACCACTT GACATTGCTC ACCACCGAGA GACTAGGGGC AAATCCGGAA CCACACTCTG AATGGATCAT CGGAAAGACG GTAGAAACT  
 9601 TCACCGTCCA CGGAGATGCC CTGGAATACA TATGGGGCAA TCACGAACCG GTAGGGTCT ATGCCAAGA GTCTGCACCA GGAGACCTC ACAGATGCC  
 9701 ACACGAAATA GTACACCAATT ACTACCATCG CCATCCCTGTG TACACCATCT TAGCCGTGCG ATCACTGCT GTGGCGATGA TGATTGGCGT AACTGTTGCA  
 9801 GCATTATGTG CTCGTAAAGC CGCGCGTGTGAG TGCGCTGACG CATATGCCCT GGCCCCAAAT CGCGCGTACCGT CAACTCCGCT CGCACTTTTG TGCTGTGTTA  
 9901 GGTCGGCTAA TGCTGAAACA TTACCGGAGA CCATGAGTTA CCTATGGTCA AACAGCCAGC CATTCTCTG GGTCAGCTG TGATACCCC TGCGCGCTGT  
 10001 CATCTCTCA ATGCCCTTGTG CTCATGCTGTTT TTAGTGGTGTG CGGGGGCTA CCTGGCGAAG GTAGACGGCT ACACATGTC GACCACTGTT  
 10101 CCAAAATGTGTC CACAGTACCC GTATAAGCA CTGGTGTGAA CGGGCAGGGTA CGCCCGCGTC AATTGGAGA TTACTCTCAT GTCCTCGAG GTTTGGCTT  
 10201 CCACCAACCA AGAGTACATC ACCTGCAAAAT TCACCACTGT GGTCCTCTCC CCTAAAGTC AATGCTGCGG CCTCTGGAA TGTCAGGGG CGCGCTACCC  
 10301 AGACTATACC TGCAAGGTCT TTGGGGGGT GTACCCCTTC' ATGTGGGGAG GAGCACAATG TTTCGGACG AGTGAAGAAC GCAAGATGAG TGAGGCGTAC  
 10401 GTCGAATTGT CACCAAGATG CGCGACTGAC CACCGCAGG CGATTAAGGT GCATACTGCC CGCATGAAAG TAGGACTACG TATAGTGTAC GGGAAACACTA  
 10501 CCAGTTCTC AGATGTGTAC GTGAAACGGAG TCAACCCAGG AACCTCTAAA GACCTGAAAG TCATAGCTGG ACCAATTCGA GCATCTTCA CACCATTCGA  
 10601 TCACAAAGTC GTTATCCATC CGGGCTGTGT GTCACACTAT GACTTCGGG AATACCGGAGC GATGAAACCA GGAGGGTTG GAGACATTCAC AGTACCTCC  
 10701 TTGACTAGCA AAGATCTCAT CGCCAGCACA GACATTAGAC TACTCAAGGC TTCCGCCAGG AACGTGTCATG TCCCTACAC CGAGGCCGA TCTGGATTCG  
 10801 AGATGTGAA AAAACAACTCA CGCCGCCAC TGCAAGAAC CGCCCTTTC CGGTGCAAGA TTGCACTGAA TCCGCTTCA CGCGTGGACT GTCATACCG  
 10901 GAAACATTCCC ATCTCTATCG ACATCCCGA CGCTGCCCTT ATCAGGACAT CAGATGCAACG ACTGGTCTCA ACAGTCAAAAT GTATGTCAG TGAGTGCAC  
 11001 TACTCAGCGG ACTTCGGCGG GATGGCTACC CTGCACTGATG TATCCGACCG CGAAGGACAA TGCCCTGTAC ATTCCGATTC GAGCACAGCA ACCCTCCAAG  
 11101 AGTCGACAGT TCATCTCTG GAGAAAGGAG CGGTGACAGT ACACCTCAGC ACCCGCGGCC CACAGCGGAA CCTTATGTA TGCGCTGTG GAAAGAAC  
 11201 AACATGCAAT CGAGAATGCA AACCCACCC TGACCATATC GTGAGGACCC CGCACAAAAA TGACCAAGAA TTCCAAGCC CGATCTCAA AACTTCATGG  
 11301 AGTTGGCTGT TTGGCTTTT CGGGGGGGCC TCGCTGCTAT TAATTTAGG ACTTATGATT TTGCTTCA GCACTGATCT GACTAGCACA CGAAGATGAC  
 11401 CGCTACGGCC CAATGACGG ACCAGCAAAA CTGAGTGTAC TTCCGGAGGA CTGAGTGTCA TAATGCACTA GGCTGGTATA TTGATCCCC CGTACCCCG  
 11501 GCGCAATATAG CAACACCAAA ACTCGACGTA TTCCGGAGGA AGCGCAGTGC ATAATGCTGC GCAGTGTGCG CAAATAATCA CTATATTAAC CATTATTTA  
 11601 GCGGACGCCA AAACCAATG TATTCGAG GAAGCATGTT GCATAATGCC ATCCACCTC TCACATAACTT TTATTTATTAA TCAACAAAAAT  
 11701 TTGTTTTTA ACATTIN

FIG. 3c

Girdwood S.A.

**A. Amino Acid Sequence of the NonStructural Polyprotein**

1 MEKPVNNVDV DPQSPVVQQL QKSFPQFEVV AQQVTPNDHA NARAPSHLAS KLELEVPTT ATTLDGSAP ARRMPSIEHQY HCVCPCMRSPE DPORMMKYAS  
 101 KLAEKACKIT NKNLHEKKD LTVLDTDPA ETPLCPHN D VTCNTRAETVS VMQDVYINAP GTIYHQAMKG VRTLYWIGFD TTQPMFSAMA GSTPATNTNW  
 201 ADEKVYLEARN IGLCSTKLSE GRTGKLSMR KKELKPGSRV YFSGVSTLYP EHRSALQSWH LPSVFHLEGK QSYTCRCDTV VSCGYVVKK ITSPGHTG  
 301 TVGYAVTNNS EGFLCKVTD TVKGEAVSPP VCTYIPATC DQMTGIMATD ISTDDAQKLL VGLNQRVNM GCTNRNTHTM QNYLLMIAQ GFSKWAKERK  
 401 EDDLNEMKLG TRECRLTYCG LWAFRTKVVH SYPRYPMOTQ TVKVPASFA PMSSVWTTS LPMSSLRKIK LALOPKKEEK LLOVPEELVM EAKAAFEDAQ  
 501 EESRAEKLRS ALPLPVADKQ IEAAAEVVCE VEGIQLADIGA ALVETPRGHV RUPQANDRM IQGYIVVSP7 SVLKNAKLAP AHPLADGVKI ITHSGISORY  
 601 AVEPYDAKVL MPAGSAPVWP EPLASERAT LYVNEKEPVN RKLYHIAHG PAKNTTEEOY KVTKAELAET EYVFDVDKKR CVKKEEASGL VLSGELTNPP  
 701 YHEALEGLR TRPVVPKVE TIGVIGAPGS GKSADIKSTV TAROLVTSQR KENCLKEIQAD VLRLRQNQIT SKTVDSVMNL GCRKAVEVLY VDEAFACHAG  
 801 ALLALIAVR PRKEVVLGDD PKQCCPFNMN QLEKVYFNHPE KDICTKTFY FISURCTQPV TAIVSLHYD GKMKETTNPCX KNIEDITGA TEPKPQDIL  
 901 TCFROWVVKQI QIDYGPHEVM TAAASQGLTR KQVYAVRKY NENPLYAITS EHVNVLTTK EDRLVWKTQ GDWPWIKQLTN YPKGQFQATI EDWEAENHKGI  
 1001 LAADNSPAPR TNPFSCXTNV CWAKRLEPL ATAGYVLTGC QWSELPQPA DDKPHSAIYA LDVICKPFG MDLTSGLPSX QSIPLTHPA DSARPVAHWG  
 1101 NSPUPTRKYG DHAVAAEBSR RPPVFLAGK CTQBLQJTG TRVSDAQHNM LVEPNKKEQI GPVKKFLSQI KIHHSVLVVB EXKIAEPHKRI  
 1201 EWIAPIAGG ADKNNYALAFG FPPQARYDLV FINIOTKYNH HNPQQCEDINA ATLETLRSA LNCLNPGOTL VVCSYQYADR NSEDVVTALA RKPVRYSAAR  
 1301 PECSVSPTEM YLIRFQLDMS RTROPTPHL NCVISSVYPT TRDGVQAAPS YRTKREMAID CQEAEVNHAA NPLGRPGIEGV CRAIYKRWPN SPTGSAETG  
 1401 TAKLTVQCQK KVHVAVGDPD RKHPEAEALK LLQNAVAYA DLVNEHNTG VAIPLLSTGI YAAGKDRLREV SLNCLTTALD RTDADVITYC LDKEWKERID  
 1501 AVLQLKESVI ELKEDMEID DELVWVHPO CLKGRKGFST TKOKLYSYTE GTKPHQAAKD MAEIKVLPN DQESNEQLCA YILGETMEA REKCPVHDNP  
 1601 SSSPPKTLPC LCMYAMPTPE VRHLRSNNVY EVTVCSSTPL PKYKIKNVOK VQCTKVVLFN PHTPAFVPAR KYIEAPEQPA APPQAEEAP EVAATTTPA  
 1701 AONTSLDVTIS ISLDMEDSSE GSLFSSFGS DNSITSMDSW SSGPSLSEI DRHQVYVADV HAQEPAPVP P9RLKXMARL AAARMQSEPT PRASTSSADP  
 1801 SIHLSPGGVS MSFGSLFDGQ MGALAAAQPP ASTCPDVPM SGFSFSDFEI EELSRYVTES EPVLFGSFEF GEVNSISSLR SVVSPFFRKQ RRRRSRUTE  
 Y

**B. Amino Acid Sequence of the Structural Polyprotein**

1 MNROFFNNMLG KRPFPAPTAM WRPRRRRQAA PMPARNGLAS QIQQLTTAVS ALVIGOATRP QTPRPRPPR QKKQAPKOPP KPKKPKTQEK KKKQPAKPK  
 101 GKRQRMALKL EADRLFDVKN EGDGVIGHAL AMEGKVKMPL HVKQTDIDHPV LEKLKPTKSS AYDMEFAQLP VNMRSEAFTY TSEHPEGFYN WHHOAVQYSG  
 201 GRFTIPKVG GRGDSGPIM DNSGRVVAIV LGGADEGITR ALSVVTWNSK GKTIKTTPEG TEWSAALPV TAMCLLGNVS FCNCRPTCTV TREPBDALDI  
 301 LEENVNHEAY DTLLTNAILRC GSRSRSRSV TDDPFLTSPY LGTCSYCHHT EPCPSPIKE QWVDEADDDNT IRQTSQAOFO YDOSOAASSN KYRYSMSLEQD  
 401 HTVKEGTMDD IKLTSGPCR RLSYKGYFLL AKCPGDSVT VSIASSNSAT SCTMARKKIP KPVGREKYL PPHGKPKC TVYDRLKETT AGYITMHUO  
 501 PHAYITSYLER SSGKVVYAKP SGKMTTYECK CGDYKTGTVT TKTETGCTA IKQCVAYKSD QTKWVFNSPD LIRHADHTAG GKHLHFFKLI PSTMVYVAH  
 601 APNHHHGPKH ISLQDLDTHL TLTTTRRLGA NEPTTEWV GKTVRNPTVD RDGLEYTWGN HEVFRVYQAB SAPGDPHGW PHEVQHYHR HPVYTTLAVA  
 701 SAAVAMMIVG TVAALACKA RRECLTYPAL APNAVFTSL ALLCCVRSAN AEITPTETMSY LW3NSQFFFV VOLCIPLAAV IVMRCCSCC LPFLVYAGAY  
 801 LAKVDAEHA TTVPVNPQIP YKALVERAGY APLMLETVW SSEVLPSTNO EYITCKPFTV VPSPKYECGG SLECPAAHA DYTCKYFGV YTFMWGOAQ  
 901 FCDSBNSQMS EAYVSEASDC ATDHAQAIKV HTAAMKVGRL IVYGNNTSFL DVYVNGVTPG TSKDLKVIAG MISASFTPD HKVVIHRLV YNYDFFBYGA  
 1001 MKPGAFGDIQ ATSLTSKDLI ASTDIRLLKP SAKNVHVYPT QAASGFEMWV NNSCRPLQET APFGCKIAVN PLRAVDCSYG NIPISIDPN AAFIKTSADP  
 1101 LVSTVKCDVS ECTTYSADFGG MATLQYVSDR EGQCPVHS STATLQESTV HVLEKGAVTY HFSTASPOAN FIVSLCGKKT TCNAECKPPA DHIVSTPHK  
 1201 DQEFLQAAISK TSWSWLFALP GCASSLLIG LMIFACSMML TSTRX

FIG. 4

## Nucleotide Sequence of S55

1 ATTGGCCCC TAGTACACAC TATTGAATCA AACAGCCGAC CAATTGACT ACCATCACAA TGAGAAGCC ACTAGTTAACG GTAGACGTAG ACCCTCAGAG TCGTTTTCG CTGCAACTGC  
 121 AAAAGAGTTT CCCCAATTG GAGGTAGTAG CACAGGGT CACTTCAAT GACCCTCTTA ATGCCAGACG ATTTTGCGAT CTGGGAGTGA AACTGATCGA CCTGGAGGTT CTACACAG  
 131 CGACGATTTT GGACATAGGAG AGGCCAECGGG TTGGTAGAAT GTTTEGGAG CACCATGACC ATTCGGTTG CCCCCTGGG ATGCCAGAGG ACCTGGACGG CATGATCAA TATGCCAGA  
 141 AACTGGCGG AAAAGCATGT AGAGTACAA ACAAGAACCTT CCATGAGAG ATCAAGGACG TCGGAGCGT ATTCGATACA CCGGATGCG AACGCCATC ATCTTGCTTC CACAAAGGAT  
 151 TTACCTGCAA CACCCGTGCG GAGTACTCGG TCACTGAGGA CCGGCTCCG GAACATTTA CCACCGGGT ATGAAAGGGC TCGGACCGT GTACTGCGAT GCGTTGCGA  
 161 CCACCCAGT CAGTTCGCG GTTATGGCG GTTGTGACCC TGATACAAAC ACCAACCTGG CCGGAGAAA AGTCTTGGAG GCGGCTAACA TCCGACCTG CACCAAGG CTCGATGAG  
 171 GCAGGACAGG AAAGTGTGCG ATAATGAGGA AGAAGGGTGT GAAAGGGGGG TGACGGTTG ATTTCTCGGT TTGAGTACAA CTTTACCGAG AACACAGG CACCTTGCG AGCTGGCGAT  
 181 TTTCATGGGT GTTCACTTG AAAGGAAAGC AGTCGATACAC TTGCGGTGT GATACTGCG TGAGCTGGGA AGCTGATCGA GTGAAGAAA TCAACATCG TCCCGGGATC ACGGGAGAA  
 191 CGCTGGATA CCCGGTACAA ACAATAGCG AGGGCTTCTT GTATGACAA GTTACCGATA CAGTAAAGG AGAACGGGTG TGTTTCCCG TGTCGACGA TATECCCCC ACCATATGG  
 201 ATCAGATGAC CGGATAATG GCCACGGATA TCTCACTGAA CGATCACAA AACATTTGG TTGGGCTCA CCAAGGATC GTTCAATACG GTAACTAA CAGGAACAC AACATACATG  
 211 AAAATTACCTT TCTGCAATG ATTCGACAA GGTTCAGGAA GACGGGAAAGG AAGATTTGCA ATGTCGGCA CCAAGAGGG CAACTTACA TATGCGCT  
 221 TGTGGCGTTT TCCGACTAAG AAAGTGTGCG CTCCTTATGG CACCGAGGCA TGCTAAAAGG CCGGAGCTCT TTGAGCTGTTT TCCCTATGTC ATCCGATGATG ACTACCTCTT  
 231 TCCCTATGTC GTTGGCGAGG AAAGTCAATG TGGTCAATCA ACCAAAGAAG GAGGAAAGC TCGTCAAGT CCGGAGGAA TTGTTTACGG AGGCAAGGG TGCTTTCGAG GATGCTCAAG  
 241 AGGAATCCAG AGCCGGAGAG TTGGAGAG CACTCCACCC ATTACTGCGA CAGAAAGGTG TGCGGAGGT GTTGGAGAG TGAGGGGGCTT CCAAGGGGAC ACGGGAGGAG  
 251 CACTCTGCA AACCCCGCGG GTTCACTGAA GGATAATACG TCAAGCAAT GACCTGAGT TGCGACAGTA TATGCGATCTT TGTCGAGAT GAGGCTAA CTCGACAC  
 261 CACACCCCTT AGCAGACATCA TAACGGCACTC CGGAGATCA GGAGGTATG CAGTCACCC ATACGACGCTT AAAGTACGTA TGCCACGGG AGTGGCGTA CCATGGCGA  
 271 AATTCTTACG ACTGAGTGA AGGCCACCGC TTGTGACAA CGAAAGAGAG TTGTGAAACG GAAAGCTGTA CCATATGGC ATGCCAGCGC CCGCTAGAA TACAGAAGG GAOCACTACA  
 281 AGGTTACAAAGG CGCAGACCTT GGAGAACAG AGTACGTTG TTGCGTGCAC AAGAACGGT CCGTAAAGG CGAACAGCC TGAGCTGCG TTCTTTCGG AGAAGTGCAC AACCCCGCT  
 291 ATCAGACACT AGCTTGTGAGG CGACTGAGA CTCGACCCGGG GTTCCCTGATC AAGGTTGAA CAAATAGGAT GAT/ JENICA CCAAGGATC CGAACGACG TATCA.....G TCAACTGTC  
 301 CGGCACTGTA TTGTTGACG ACCCGGAGA AGAAAGACCTT CGGCAAAATTG GAGGGGGCGT TGTCACGGT CAGATCACTG CCAAGACAGT GGATTCGTTT ATGTCACAG  
 311 GATOCACAA AGCGGTAGAA GTGCTGTGATG TTGACGAGG GTTGGCGGT CACGGAGGAG CACTACTTCG TTGTTGCGA ATGTCACAGG CCGGCTAGAA GGTAGTACTA TGCGGAGAC  
 321 CTAAAGCAAGG CGGATTTCTC AACATGATG ACCTAAAGGT ATCTACACG CACCTGAAAGA AAGACATGATG TACCAAGACG TTATCACTG CAGCTGGCG AGCTGGCG  
 331 CGGTTATGTT ATCCGACTG CATTACCGATG GAAACACACG CCACTGAAAGA AGAACATGCA ATGACGATC ACAGGGCGA CGAACGGGA CCAOOGGGAC ATCACTGCG  
 341 CATGTTTCCG CGGTTGGGTTT AAGCAACTG AAATGCACTA TCCGGACAT GAGGTAATG CACGGGGGGG CTACACAGAA AAGGAGTATA TGCCGTTGGG CAAAAGGTA  
 351 ATGAAAACCC GTGTGACGG ATCACATGAG AGCATGTGAA CTGTTTCTC ACCCCGACG AGGACGACGT AGTATGAAAGG ACTTTACGG CCGGACCTATG GATTAACGAG CTCACATGAG  
 361 TACCTAAAGG AAATTTCAGG CGCACCACCG AGGACTGGGA AGCTTAAACG AAGGGATAAA TTGCTCGAT AAGACGTTCTC GTTCCCGGTA CCAATGCGTT CAGTCGAAAG AGTACGTT  
 371 GTTGGGGGAA AGGACTGGAA CGGATACTGGG CCAAGGGGGG TATGCTGTT ACCTGGTTGTT AGTGGAGGA GTGTTTCCCA CATTGTTGGG ATGACAAACG ACACTGGCC ATCTACGCT  
 381 TAGACGATAT TCTGCAATG TTGTTGGCA TGAGCTGGT AAGGGGGGTTT TCTGAAAGG AGGACGATCCG GTTAAAGTGC CTCCTGCGG ACTACGCGG GCGAGTACG TATGGGAGA  
 391 ACAGGCGGAGG AACACCGGAGG TATGGTACG ATCACCGCTG TTGGGGGAA GTTCCCGGTA GATTCGGCGT GTTCCAGGTA GCAACAGCG AGGGGGAGG CGGGCGAGAA  
 401 CTAGAGTTAT CTGTCGACG CATAACTGCG TCCGGTCAA CGGCAACTC CTCACGGCTT CTGTTGGCGA CGCAAGGAG AACAAACCCG GCGGGCTGCA AAAATTCTT AGGGAGTCA  
 411 AACACCCATC CTGTTGCTG AGTCTGAGA AAAAATGAG AGCTTCCACG AACGAAAGT AATGCTGCG CCGGATGCG ATGACGATGCA CGCGGTTGCG CTGTTGCTC  
 421 TTCCCCCGCA CGGACGGTAC GACCTGGGTG TCTACATATAA TACAGAAGG ATCACTTCA ACAGTGTGAGG GACCCGGCG CGACCTGAA AACCCCTTGGG CTGTTGGCC  
 431 TGAACGTTCT TAACCCCGA CGGACCTGGG TTGGTGAAGTC CTACGGTTAC CGGCAACCGA ATAGTGAAGA CGTAGTCACG GTTCTTCCCA GAAAATTGTTT CAGAGTGTGCG CGAGGAGGCG  
 441 CAGAGTGGT CTCAAGGAT AAGGAAATGT AGCTGATTTT CGGACAACTA GACAACAGG CCAACGGACA ATTCACCCCGT CACATTGAGA ATGTTGCGAT TTGCTGGGT TACGGGGGT  
 451 CAAAGGAGG AGTGGAGCG CGACCGCTGT AGGCTACTAA AAGGGAGAAC ATGGCTGTT GTCAAGAGGA ACCAGTGTG ATGCAAGGCA ATCCACGCG CAGGACGG  
 461 CGGCTGCGAT CTATAAGCTG TTGCCCCGAA GTTCCACCGA TCAACGAGGTA CACGGAGGTA CGGCAAAACTG GACTGTTGCG CACGGAGAGA AAGTGTACCA CGCGGTTGCG CTGTTGCTC  
 471 GGAACACCC AGAGGCGAGG CGGCAAAATG TCTGGTAAAGG CGGCTACCAT CGGAGTGGCG ATGTTGAGGTT TCAACATAAT ATCAAGTGTG TGCCGATCC AGCTGATGCG ACAGGCGAT  
 481 AGGGAGGGG AAAAGACCGG TTGAGGTTG CACTTAAACG CTGGACAGA GACGATGACG TTGAGGTTG CACGGTACCC TCTGTTGAGGTT TCAACATAAT ATCAAGTGTG TGCCGATCC AGCTGATGCG  
 491 CGGCTGCGAT CTGAGGAGG AGATGAGGAG GTTGGAGGAG GAGGGGGGG ACCGGGGGGT ATGGCTGATC CGGCAAGCTG GTTCCAGGAGG AAGGAGGGG TTGAGTACTA  
 501 CAAAGGGAA GTTGTATGTT TACTTTGAG CGGCAAAATTG CGTACGAGCA CGGAAAGATA TGCCGGAGAT AAGGCTGCG TTGCTTAAAGG AGGAGGAAACG CACGAAACAA CTGTTGCG  
 511 ACATATTGGG CGGAGGAGG GAGGCAATCC CGGAAAGGTTT CGGCTGCGAG CACAAACGGT GTTCTGAGCC CGGAAAGGCG TTGCTGGCG TTGTTGAGG TGCCATGAGG CGGAGAAGGG  
 521 TCCACAGCT CAGAAACAT AACCTCAAG AGTGTACGT ATGCTCTCC ACCECCCTTC CAAAGTACAA AACGAAAGT GTTCAAGAGG TTGAGTGGC AAAAGTAGTC CTGTTAACC  
 531 CGGCTACCC CGGATTTGGT CGGCCCCGTA AGTACATGAGG AGGAAAGGAGG CACGGGGGGG CGGGGGGGGG GAGTTTGGGG GACACCAACA CGACCTGCG  
 541 CTGATAACAC CTGCTGTTG GTTCAAGGAGA TCTACGAGA CTTGAGGAGG AGTGGAGGAGG GCTACCTGTTT AGGGGATCCG ACAAATACCG AAGGGAGGTT GTGCTGGCG  
 551 AGCTCCATG CGTGGCGGAGG CTOGCCCCGGG TCTACGAGGAGG AGGCTAAAGG AAGGCTGCG GCTGGGGAG GCAAGAAGT GAGGAGGAGG CACCTACCCG CGGCTACCC  
 561 AGGAGTCCCT TCACTTCTT TTGATGGGGG TATGCTATGCTT CTGGGAGGAGG GAGGAGGAGG CGGCAACAGG CGGGGGGGAG TACATGCGCTT AGGGCTGCG  
 571 CTATGCTTCTT CGGATGCGTTT CGGAGGGAGG AGATGAGGAGG GTTGGAGGGGG AGAGTGGGGG GCTCTGTTT GGGCTGTTT GGGCTGTTT GGGCTGTTT GGGCTGTTT  
 581 ACTTCCAAAAGA AGAGTGGGGT GTTCAAGGAGG AGCTTACAGA ACCGACCTGG GAGGGCAATG TTGTTGAGG AATCTACCCG CGGGGGGGGG AGGAGGAGG CTCACATGCA  
 591 GTTACCCATG GATCCCCACCG GAGGCAACAA AACGAGGAGG CACGGCTGCA AGAGTACAA ACCGAGAACG CATAACACCT CGGAGGACTCC TTGCAACGCTT AGGGCTGCG  
 601 CAGATGACCC AGAAATGCTAT AGGCTACCTT ACCGGAAACC ATGCTTACG AGGCTGATAC CGGGCAACTA CTGTTGAGCA AAGGTTGCGT TAGGTTGTTG TAAACAACTAT CTGATGAGA  
 611 ATTACCGACG CGTACGATCT TATGAGATCA CGGAGGAGG AGTGGGGAGG AGTGGGGAGG CTGAGTACCTG TACGGGGAGG AGTGGGGAGG CTGAGTACCTG CACCTTGGG  
 621 ACCEGAAAG AGACGGAGT AGAGGCCCCAA AGCTTCCCGAG TOGGTTGCA CGAGGAGATG AGAACACGTT CGAAAGCTG TTGCTGCGTT CGCTACGAGG  
 631 TGCGTGAAGT CGCAACACTG GACCTACGGG CATTCAACGTT TGAATGCTT CGGAAATGAG TCAAGGTTG GAGGAGTGG CGGAGGAGG AACCTGGGG  
 641 TGCTTACCCGCTT ATAGTGGCC AGACTGAGGAG CGCTTACGGG CGGGGGGGT TTGCAAGAGA CGGCAAAATTG GTTCCCATGAG CAGAGGAGG CTAGTACGAG ATTCGTCATG  
 651 GAGACGTTAG AGTTACACCT CGCACGAAAC ACACAGAAGG AAGACGGAAAGG CTACAAGTGAG TACAAACCCG AGAACCCCTG CGGACCGCTT ACCTATGCGG GAGTTAGTGC

Fig 5A

6721 GCAGGCTTAC AGCCGGTTTC CTACCCAAAC TTCACACGGCT CTTTGACATO TCGGGCGAGG ACTTTGATGC AATCATAGCA GAAACTTCA ACCAAGGTGA CCCGGTACTG GAGACCGATA  
 6841 TCGGCTGGTT CGACAAAAGC CAAGACGGCG CTATGGCTT AAACGGCGTAA ATGATCTTGC AAGACCTGG TGTGACCAA CCACACTCG ACTTGATGA GTGGCCCTTT CGAGAAATAT  
 6981 CATCCACCA TCTGCCAACG GTTACCCGGT TCAATTCTGG CGCGATGATG AAATCCGAA TGTTCCTCAC CCTTCTTCAC AACACAGTC TGAATGCTGT TATGCCAGC AGATGATGG  
 7031 AGACGGCGCT TAAAGCTTC AAATTCGAG CATTATCTGG CGACGACAC ATTATACAGC GAGTAGTATC TGACAAGAA ATGCTTGAGA GGTGCTCAAC ATGGAGGTTA  
 7201 AGATCTTGA CGCAGTCATC CGCAGAGAC CACCTACTT CTGGGTGGA TICATCTTC AAGATGGGT TACCTECACA CGTGTGCGG TCGGGGACCC CTGGAAAAGG CTGTTTAAGT  
 7321 TGGTAAACCC GTCTCCAGGC GACGATGACG AACACAAAGC CAGAAGACCC GTCTCTGAGT ATGAAACAA CGCGTGGTTT AGAGTAGGTA TAACAGACAC TTACAGCTG GGTGCGGCAA  
 7441 CTGGTATGA GGAGACAC ACACACCTG TCTGCTGGC ATTGAGAACT TTGACCAAGC GCAAAAGGC ATTCAAGGC ATCAGAGGG AAATAAAGC TCTCTAGGT GGTGCGGAA  
 7561 AGTCAGGATA GTACATTCA TCTGACTAAAT ACCACAAAC CACCAACATG AATAGAGGT TCTTAAACAT GTCTGGCGC CGCCCTTCAC CAGCCCCAC TCCATGTTG AGGGGGCGGA  
 7681 GAAGGAGCA CGCGGGCGG ATGCGCTCCC GCAATGGGT GGTTCCTCA ATCCAGCAAC TCAACACAGC CTGACTGTC CTAGTCATG GACAGGCAAC TAGACCTCAA ACGCCACGCC  
 7821 CACGGCTCCC CGCCGGCGAG AAGAACGAG CGCGAAAGCA ACCACCGAG CGAAGAAC CAAACACACA CGAGAACAG AAGAACGAACT CGCCAAACCC CAAACCGCGA AAGAGACAG  
 7921 GTATGGCACT TAAGTGGAG CGCGACAGC TTCTGGAGT CAAATGAG GACGGAGATC TCACTGGCA CGACATGGCA ATGAAAGGAA AGGTAAATGAA ACCACTCCAC GTGAAAGGAA  
 8041 CTATGACCA CGCTCTCTCA TCAAGCTCA AATTCACCA GGTGCTGAGA TACGACATC AGTGGACACA GTTGGGGTC AACATAGGA GTGAGGGTT CACCTACACC AGTGAACACCC  
 8161 CTGAAAGGTT CTACACCTGG CGACGGAGG CGTGGCTAAT TGTGCTGG AGATTTACCA TCCCGGGGG AGTAGGAGGC AGGAGGAGA GTGGTGTGTT GATTATGGAT AACTCAGGCC  
 8281 CGGTGTTGCG GATAGCTCTC CGAGGGCGT ATGAGGAAAC AAGAACGCC CTGGTGGTC TCACTGGAA TACGAAAGG AACAAATCA AGACAAECCC CGAAGGGACA GAAGAGTGT  
 8401 CTGCTCACCC ACTGGTCAAG CGCTGATTCCTC TCTCTGGAAA CGTGGCTTC CCAATGCACT CGCCGGGGCAC ATTCACAC CGAACGGCT CGACATCTTC GAAGAGAACCC  
 8521 TGAACCAAG CGCGTACGAC ACCCTCTCTCA ACCGGCATATT CGCGTGGCGA TGTGGCGCA GAAGTAAAG AGGCTGACT GAGGACTTCA CTGTCACCG CGCGTACTG CGCACATGCT  
 8641 CGTACTGTC CGATACGAA CGTGTCTTAA GACGACGGC GTTGGGATG AAGCGACGA CGACACATA CGACATAGA CTTCGGCGA GTTGGGATAC GACCAAAAGC  
 8761 GACCGCAAG CTAAATAAG TACCGCTACA TGTGCTGCA GAGGATCAT ACTGTCAAAG AAGGACCAT GGTGACATC AAGTCACCA CCTTCAGGCC GTTGTAAAGG CTGAGCTTCA  
 8881 AAGGATCTT TCTCTCTGG AAGTGTCTC CGGGGGACAC CGTAAACGTT ACCATGCGCA GTGACACTC ACCAAGCTCA TCGACAACTG CGCGCAAGT AAAACCAAAA TTGGTGGCAC  
 9001 CGGAAAAATA TGACCTACCT CGCTTCTGAG GTAAAGAGT TCTTTCACCA GTGACGGTC GTTGTGAAAG AACAAACCCCG CGTACATCA CGTACACAG CGCGGGACCC CACCGCTTAA  
 9121 CATGCTATCT CGGAAAGTCA " " TGGAAAGG TTACGGGAA CGCACCTCA CGGAAGAAC TACGACAGA CGTACATCA AGGCGGAA CGTACACAG CGTACACAG  
 9241 TCACCGCTTAC CACCGCTG CGACGGCTG TCCCTATAA GACGACGGC CGGAGGTTAC TCTTCAGGAT CGTACACAG CGTACATCA AGGCGGAA CGTACACAG CGTACACAG  
 9361 TCCCTTCAAC CGTGTCTGG AGTACCTCA TGTGCTGTG TGCCTACCGG CGGAACGGTAC TACCGGGTT TAAACACATE ACCCTACAT GACGACAGA CGTACACAG CGTACACAG  
 9481 CGAGGAGCT AGGGGGAAA CGGGAACCA CGACGAACT GATCATGGG AACACGGTAA GAAATCTAC ACCGGTACCGA GTGGCTGG AATACATATA CGGCAATCAC GAAGGAGTAA  
 9601 GGGCTATGCC CGAACAGAG ACCTCTACGG ATGCCACAC GAAATAGTAC ACCGTTACTA TCACTGGCAT CGTGTGACA CGATCTTAC CGTACATCA CGTACACAG CGTACACAG  
 9721 CGATGATGAT TGGTAACT GTTGGCAT TATGCTGTT TAAACGGCG CGTGTGTC TCACTGGCA CGAACGGCG CGTACACAG CGTACACAG CGTACACAG CGTACACAG  
 9841 GTGTTAGCT CGCTAACTCA GAAACATTCA CGGACGACAT GAGGTTACTA TGTGCAACA CGCAGGGTT CTCTGGGTC CGCTTGTGTC CGTGTGTC CGTGTGTC  
 9961 CGTGTGTC  
 10031 TTGAAAGGG AGGGTACGCC CGCGTCATT TGAGGATTC TGTATGTC CGGAGGGTT CGTCTTCAC CACCAAGG TACATACCT CGAACATTC ACCTGGTGC CGCTTCTCCCA  
 10241 AAGTCAGATO CTGGCTCTC TTGGATGTC ACCGGGGCGC TCAACGAC TAACTCTCA CGGTGTTGGT CGGAGGTTGG ACCGTTGTTG CGCTCTGGT CGTACACAG  
 10321 AGAACAGCA GATAGTGG CGTACGGTAC ATTTCAGT AGATGGGG CGACGGACCG CGGAGGGAT TAAAGTGCAT ACTGGGGCA CGAACGGGG TGAAGTGGG ACTGGTATA CGTACGGG  
 10441 ACACATACCG TTTCCTAGAT GTGTAAGCT ACCTGGACCG ACCGGACGG TCAAAAGACCG TGAAGTGCAT ACCTGGACCA ATTTCAGTAT GTTGTACAC CGGAGGGTTA  
 10561 TCAATGGCGG CGTGGTGTAC AACTATGACT TTCCGGATAA CGGAGGGAT AAACCGAGG CGTGGGAGA CGTCACACT ACCCTCTGA CTGACAAAGA CCTCATGCC ACCACAGACA  
 10681 TTAGGCTACT CAAGCTTCC CGCAAGAACG TCACTCTCC GTACACGAC CGCCGATCTG GATGGAGAT GTGGAAAAC AACCTGGCC CGCCACTCA CGAACACCC CGTGTGTC  
 10891 CGAACATTCG AGTCAATCTG CTGGACGGG TGCACTGTC ATACGGAAAC ATTCCTTAT CTATGACAT CGGAAACCT CGCTTATCA GGACATCA CGTACACAG CGTACACAG  
 10971 TCAAAATGCA TGTCACTTAA CGGGGGACT CGGAGGGATG CGTACCTCG AGTGTGTC CGACGGGGAA CGAACATGCC CGTACACAG CGTACACAG CGTACACAG  
 11041 TCCAAGAGTC GACAGTCAT GTCTGGAGA AAGGGGGT GACAGTACAC TCACTGGAC CGAGGGCA CGGACCCACA CGGAAACTTC ATTATATGCC TGTGTGTTAA GAAGACAAACA CGAACATCG  
 11161 AATCCAAACCC ACCGGTGTAT CATACTGCA CGACGGGGCA CAAAGATGAC CGAACATTC CGGGGGCAT CTAAAGTAC CGTACACAG CGTACACAG CGTACACAG CGTACACAG  
 11281 CGCTTAAAT TATGACCTT ATGATTTTG CTGGACGAT GATGGTAC ACCGACACGA GATGACGGGT CGGGGGCAT CGAACACCCG CGAACACCCG ATGATCTTC CGAACAGTCA  
 11401 TGTGCTATAAT CGATCAGGCT GTATGATTTAG ATCCCGCTT CGGGGGCA ATATGCAAC ACCAACACCT GAGCTTATTC CGGAGGAGGG CAGTCATAA CGTACACAG CGTACACAG  
 11521 TAATCACTAT ATTAACCAATT TATCAGGG CGCCCAAAAC TCAATGATT TGTGAGGAAG CGTGTGTC ATGACCATGC AGGGCTGCA TAATTTTTA TATTTCTTT TATTAATCA  
 11641 CAAATTTTG TTTTAACTC TTC

FIG. 5 B

## Nucleotide Sequence of TR339

1 ATTCGGGG TAGTACACAC TATTGAATCA AACAGCGAC CAATGCACT ACCATCACAA TGGAGAAGCC AGTAGTAAAC GTAGACGTGAC ACCCCCCAGG TCGGTGTTGC GTGCAACTGC  
2 AAAAAAAGCTT CCCGCAATTG GAGGTAGTAG CACACCGGT CACTCCAAAT GACCACGCTA ATGCCAGACG ATTTCGGCAT CTGGCCAGTA ACTAATGCA GTGCGAGGT CCTTACCAAGCAG  
3 CGGAGATCTT CCACATAGGC AGCCGACCCG CTGGTAGATA GTTTCGGAG CACCAAGTAC ATTGTGTTCTG CCCCCATGCGT AGTCCAGAAG ACCCGGACCG CAGTGTGAAAT TATGCCAGA  
4 AACTGGGG AAAGGCTGC AGGATACAA ACAAGAACCTT OCATGAGAAG ATTAAGGATC TCCGGACCGT ACTTGATACO CGGATGCTG AAACACCATC GTCTGCTT CACAAGGATC  
5 TAATCCCAA CATGCTCC GAAATTCGC TCATCAGGA COTUTATATE AACGGCTECCG GAATCTATCA TCACTACGGT ATGAAAGGCG TCGGACCCCT GTACTGATTG CGCTTCGAC  
6 CCACCGAGT CTATGTTCTG GTCTGGACAC AACAACTGGG CGGACGAGAAG ATGCTGTTGAA CGCCGATTAAC TCGGACTTGT CACGACAAAG CTGAGTGGAAAG  
7 GTGAGCAGG AAAATTGTCG ATAATGAGGA AGAAGGAGTT GAAGCCCGGG TGCGGCGTTT ATTTCGGTGT AGGATGCGA CTTTACCGA AACACAGGA CACCTCCGAG AGCTGGCGATC  
8 TTTCATCGGT GTTCCACTTG AATGAAAGC AGTCUACAC TGGCGGTGTG GATACAGTGG TGAAGTGGCA AGGCTACATA GTGAGAAGAA TCAACATCG AGCTGGCGATC  
9 CGTGGGATA CGGGGTATCA CACAATACCG AGGGCTTCTT GTTACATCAA GTTACGACAA CAGTAAAGG AGAACCGGTA TCGTCCCTG TGTGACCAAGT CACCCCGGGC ACACATAC  
10 ATCAGATGAC TGTGTTATAATG CGCACGGATA TATCACCTGA CGATOCACAA AAACCTGGT TCGGCTCTAA CGGCGAATT GTCTTAAACG GTGACGAA CAGGACACCC AACACCATG  
11 AAAATTACCT TCTGCGGATC ATAGCACAAAG GOTTAGCAA ATGGGCTAAG GACGCCAAGG ATGATCTGAA TAAAGGAGAA ATGCTGTTA CTAGAGAACG CAACTTACG TATGGCTG  
12 TUTGGGCGTT TCGGACTAAG AAAGTACATT CTTCATGCA CGCACCGTGA ACCGAGACCA CCGAGCTTGT TGTAAAGGCG TGTACCGCTT TGTGACCGT GTCTGCTTACG AGGCGCTTGT  
13 TCCCGATCGT GCTGGGGAGC AAATGGAACG TGTGATGCA ACCAAGAAGG GAGGAAAGAAC TCTGGCGAGT CTGGACGAA TTAATCAGG AGGCGAACAGC TGTGTTGCG AGTGTGCG  
14 TCCCGATCGT GCTGGGGAGC AAATGGAACG TGTGATGCA ACCAAGAAGG GAGGAAAGAAC TCTGGCGAGT CTGGACGAA TTAATCAGG AGGCGAACAGC TGTGTTGCG AGTGTGCG  
15 AGGAAGCGG AGGGAGAGG CTGGAGAAG CACTTCACCC ATTAGTCCA GACAAAGCA TCGGGCGAC CGGAGAAGTT GTCTGGAGAAGT CGGAGGCGAC ATCGGGAGCA  
16 CATTAGTTG AAAAAACCGGG CGTCAGCTAA GGATATAACG TCAAGGAAAT GACCGATGAA TCGGAGACTA TATGGTGTG TGCGGAAACT CTGTTGTTA GAATGCGAA CTGGAC  
17 CGCACCGGGT ACCAGATGAC GTTAAAGATA TAAACACACT CGTATGAGA AGGAGGATCC CGGTGGAAACG ATGAGCGCT AGAAGTCTG TGCGGACCG AGGCGAC  
18 AATTCCTACG ACTGAGTGAG AGCGGCEACGT TAGTGTACAA CGAAAGAGG TTGTGAAACG CGAAACATACT CGCAGATGCG ATGAGTGTGCG CGGAGGCG  
19 AGGTTACAAA CGCAGAGCTT CGAGAACAG AGTACGTGTT TGACGTTGAC AAGAAGGGT GTGTTAAAGG CGGAGAAGCC TCAACATGCA TACAGAACG GAGCAGTAC  
20 ATCATGAGCT AGCTCTGGAG GAGCTGAAGA CGCCGACCTGC GGTGGCTGAC AAAGCTGAAAGA CAATAGGAT GATAAGGACA CGGGGGTGG CGAAGTCCG TCAACATGCTG  
21 CGGACCGGG AATGGGGAAAGG AAAAATGGT CGGCGGAAATG TGCGGAGCT CGTGGCGTGC CACCGAGGAG CACTACTTGC CTGGATGCTG ATGCGGAGCT CGGAGGCG  
22 GATGCGGACAA AGCCGGTAGAA TGCGGATCG TGCGGAGCT CGTGGCGTGC CACCGAGGAG CACTACTTGC CTGGATGCTG ATGCGGAGCT CGGAGGCG  
23 CGTACATGCGT CGGAGGAGG AAAAATGGT GATGCGGAGCT CGTGGCGTGC CACCGAGGAG CACTACTTGC CTGGATGCTG ATGCGGAGCT CGGAGGCG  
24 CAGCTATGTT CGTACATGCGT CGGAGGAGG AAAAATGGT GATGCGGAGCT CGTGGCGTGC CACCGAGGAG CACTACTTGC CTGGATGCTG ATGCGGAGCT CGGAGGCG  
25 CATGTTCCG CGGGTGGGT AGGCAATTG AAATGCGTAA TCCGGGACAT GAAGTATGCA CGGGCGGCG CTACAAAGGG CTAAACAGGA AGGAGGTTA TGCGGCGG CAAAAGTCA  
26 ATGAAAACCC ACTGATGCGG ATCACATGAG AGCATGTTAA CGTGTGTTCT ACCGGGACTG AGGAGCGCT AGTGTGAAA ACCTTGGAGG CGGACCCCGT GATTAAGCG CTCACAAAC  
27 TACCTAAAGG AAACCTTACG GTGACTATAG AGGACTGTTGA AGCTGAAACG AAGGGAATAAA TGCTGTTAA ACACGGCCC ACCTCCCGT CGGACCCCGT GATTAAGCG  
28 GCTGGGCGAA CGGATTGGAA CGGATACTAG CGACGGGGG TGATGTTACTT ACCGGTGCCTG AGTGGAGCGA CGGGTGGCG ATGACAAGG ACATTCGGC ATTTACCGCT  
29 TACGATGATGAA TGTGATGAA CGGATACTAG CGACGGGGG TGATGTTACTT ACCGGTGCCTG AGTGGAGCGA CGGGTGGCG ATGACAAGG ACATTCGGC ATTTACCGCT  
30 TACGATGATGAA TGTGATGAA CGGATACTAG CGACGGGGG TGATGTTACTT ACCGGTGCCTG AGTGGAGCGA CGGGTGGCG ATGACAAGG ACATTCGGC ATTTACCGCT  
31 GCTGGGCGAA CGGATTGGAA CGGATACTAG CGACGGGGG TGATGTTACTT ACCGGTGCCTG AGTGGAGCGA CGGGTGGCG ATGACAAGG ACATTCGGC ATTTACCGCT  
32 TACGATGATGAA TGTGATGAA CGGATACTAG CGACGGGGG TGATGTTACTT ACCGGTGCCTG AGTGGAGCGA CGGGTGGCG ATGACAAGG ACATTCGGC ATTTACCGCT  
33 ACAGCCCGAGG AAACCCCGAA TGTGATGAA CGGATACTAG CGACGGGGG TGATGTTACTT ACCGGTGCCTG AGTGGAGCGA CGGGTGGCG ATGACAAGG  
34 CGGAGGTTCT CGTCCACAGG TCAACATGCG TGGCGGTTAA CGGCAATTTT CTGTCACGGT TAGTGTGCGG TGACAAAGGG AACCCGGGGG CGGGGGGGG  
35 AACACACTTCG ACTGAGTGTG TGCGGAGCT CGTGGCGTGC CACCGAGGAG CACTACTTGC CTGGATGCTG ATGCGGAGCT CGGAGGCG  
36 TITCCCGGCGA CGGACCGGGT GACCTGGGTTGT CTACATGCGT CGGAGGAGG AACCTTGGCG ATGCGGAGCT CGGAGGCG  
37 TGAATTGCT TAACCCAGG CGGACCCCTCG TGTGAGGTC CTATGGCTAC CGGGGACCGA AGCTGAGGAG CGTACGTCACG CTGTTGCGCA GAAGGTTGTT CGGGGGCG  
38 CAGATGTTGT CTCAAGCAAT ACAGAAATGT AGCTGATTTT CGGACAACAA GACAACAGCC GTACACGGCA ATTCACCCCG CACCATCTGA ATTGCGTGT TGCGGCG  
39 CAAGAGATGG AGCTGGAGCC CGGGCGCTAT ACCGGACCAA AAGGGGAAAT ATTGCGTGT GTCAAGAGGA AGGAGGAGCA ATTCACCCCG CACCATCTGA ATGCGGAGCG  
40 CGGGTGGCGAT CTATAAACGTT TGCGGAGCCA GTTGTACCGG TCAACCCAGG GAGCAGAGCA CGGGGAAAGT GACTGTTGTC CTAGGAAAGA AGTGTGCA CGGGGGCG  
41 CGGAGACCGE AGAAGGAGAA CGGTTGAAAT TGCTACAAAA CGGCTACCAT CGTGGAGCG ACTTAGTAAAG TGAACTACG ATCAAGTGTG TCGGCGATTC ACAGGCGATT  
42 AGCGGAGCCG AAAAGGCGCC CTGAGAGTGT CACTAATGCG CTGAGAAACG CGGCTGAGCA GAAGTGGAGCG CGACGTAACG ATCTATGCG TGCGGAGGAA AGAATGCG  
43 CGGCGATCTCA ACTTAAAGGG TGCTGAGAGG AGCTGAGAGG TGAGATGAG GAGATGCG AGTGGTGTG AGTGGTGTG AGTGGTGTG AGAAGGGG TTGAGTACTA  
44 CAAAGGAAAT ATGTTGAGG TACTTGGAG CGGACCAATT CGTACACAA CGGAGGAGCA CGGGGGGGAG ATGGGCTCG TTGCTGTTAG CGGAGGAGG ATGCGGAGCT  
45 ACATATGGG TGAGACATCG GAGGAACTCG CGGGGGGGG CATAACCCCG CGTCTGACCC CGGCGGGGGG CGGGGGGGG CGGGGGGG  
46 TCCACAGACT TAGAGGAAAT AACGCTGAAAG AGTGTACAGT ATGCTCTCC ACCTCCCTTC CTAAGCACAA AATTAAGAAT GTTCAAGAGG TCACTGTCAC GAAAGTGTG  
47 CGGACACATTC CGCATTGCTT CGGGGGGGGAA AGTACATAGA ATGCGGAGAA CGGCTACCG CGTCTGCTCG ACAGGGGGAG GAGGGGGGGG AGGTGTTAG GACACCGCTCA CGATCTACAG  
48 CTGATAACG CTGGCGTGT GTACAGAGCA TCTCTACGGA TATGGATGAC AGTACGGAAG GTCACGTTT TCCGAGCTT ACGGGATCGG ACACATCTAT TACTAGTATG GACAGTGTG  
49 CGTCCAGGACG TACTTACGAA GAGTGTGAG CGGGGGGGG CGGGGGGGG CGTGGCGTGT ATGCGGCTCC ACAGGGCTGG CGGGGGGG  
50 CGGGGGGGGG AAAAGGCGCC ATGCGGAGCA CGGGGGGGG CGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG  
51 CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG  
52 CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG  
53 CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG  
54 GATCATTGTA CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG  
55 TAGGCGGTA CATATTTGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG  
56 CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG  
57 CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG  
58 CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG  
59 CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG  
60 AGGGGAGCTGCTGCA CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG  
61 AGTGGCTGTG AGCTGTTGTTG ATCAACATGAC TCTACGGAAGA CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG  
62 CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG  
63 TCTGATGCGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG  
64 AGGAGTGGCG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG  
65 AAGAAGTGGCG TGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG CGGGGGGGGG

Fig 6A.

6721 CGACTGTTA CTTATCGGG ATTACCCCCG AATTAGTCCG TAGCCTTACG GCGCTTTCG TTCCAACAT TCACAGCTT TTTGACATGT CGCGCGAGA TTTTGATGCA ATCATAGCG  
 6841 AACACTTCAA GCAAGGGAC CGGGTACTGG AGACGGATAT CGCATATTC GACAAAGCC AAGACGACCG TATGCCGTA ACCGCTGCA TGATCTGGG CGACCTGGT GTGATGATCAC  
 6961 CACTACTCGA CTTGATCGAG TCGGCTTGT GAGAAATATC ATCCACCATG CTACCTACCG CTACTCGTT TAAATTCGGG GGATGATGAA ATACCGAAT TTGCTCLAA CTTTTUTCA  
 7081 ACACAGTTT GAATGTCGTG ATCCGAGCA GAGTACTAGA AGACGGCTT AAAAGCTCA GATOTOCACG GTTCACTGGC GAGGACAACG TCATACATGG AGTAGTATCT GACAAAGAA  
 7201 TGGCTGAGG GTGGCGCCACCG TGGCTCAACA TGGAGGTTAA GATECATGC ACAGTCATCG GTGAGGACG ACCTTACTTC TCCCGCGAT TTATCTGCA AGATTCGTT ACTTCCACG  
 7221 CGTCCCGCTG CGCGACCCG CTGAAAAGGC TTGTTAAGTT GGGTAAACCG CTCCCAACCG ACAGGAGGAG ACAGGAAGAC AGAACACCGG CTCTGTTAGA TGAAACAAAG CGTGTGTTA  
 7441 GAGTAAGTAT AACAGGACAT TTAGCAGTGG CGGTGACGAC CGGTGATGAG ATGACATAA TTACACCTT CCTACTGCA TTGAGAAGCTT TTGCGAGG CAAAGAGCA TTCCAAGCA  
 7561 TCAGAGGGGA AATAAAAGCAT CTCTACGGT GTCCTAAATA ATGACATGAC TACATTTCTG CTGACTAATA CTACAACACC ACCACCATGA ATACAGGATT TTAAACATG CTGGCCGCC  
 7681 CGCCCTTCCC CGCCCGCCACT CGCATUTGGA CGCCGGGGAG AAGGAGGAG CGGGCGCGA TGGCTCGGGG CAACGGCTTCTGCTCAA TCCACAACT GACCCAGCGG GTCACTGCGC  
 7801 TACTGATGG ACAGGCACT AGACCTCAAC CCCACGCTCC ACCECCCGCA CGCCGGCGAGA AGAAAGCGCG CGCCCAAGGCG CCACGGAAAGC CGAAAGAAACC AAAACCGAG GAGAAGAAGA  
 7921 AGAAGAACCC CGCAAAACCC AAACCCGGAA AGAGACACCG CATGGCACTT AGTTGAGG CGGACAGTGG GTGGACGTC AGAAAGGAG ACCGGATGT CATCGGCGAC CGACTGCGA  
 8041 TGGAAAGAAA GTTAATGAAA CTCTGACG TGAAAGGAAAC CATGACCAAC CGCTGTTCTGAAAGCTTAAAGCTTAA ATTACACAG TTGTGACGAT AGGACATGGG GTTCCGACAG TTGCGAGTCA  
 8161 ACATGAGAG TGAACCATCC ACCTACACCA GTGACACCC CGAACGATTC TATAACTGG ACCACGGACG GTGCAAGTAT AGTGAAGGTA GATTGACAT CGCTCCCGA GTAGGAGCG  
 8281 GAGGAGACAG CGTGTGTTCC ATCATGGATA ACTCCGGTGG CGTTGCTGGG ATAGTCTGCG GTGGAGGTA TGAAAGGAA CGAACTGCCC TTGCGTGTG CACCTGAAAT AGTAAAGGGA  
 8401 AGACAATTAA GAGGACCCCG GAAGGGAGC AGAGTGTTC CGAACACCA CGTGTGACG CGAATGTTT CGCTGAAAT ATGAGCTTCC CATGGACCG CGCCCGCCACG TGCTATACCC  
 8521 CGGAACTTC CAGACCGCTT GACATCTTAA AGAGACAGT GAACCATGA CGCTGAGATA CGCTGTTCAAG TOCCATATG CGGTGGGGAT CGTGTGCGAG AAGCAAAGA AGCGTCACTG  
 8641 ACCGTTTACG CTGACCCAGC CGCTTACGGT GCACTACCTG CTACTCCAC CATACTAAC CGTGTGCGAC CGCTTGTGAG ATGAGGAGG TGTGGACGA AGCGGACGAT AACACCATAC  
 8761 CGATACAGAC TCCCGCGAC TTTGGATACG CGAAAGGGG AGGACCAAGC CGAAACAGT ACCGCTACAT GTGCTTGGG CAGGATCAAC CGGTTAAAGA AGGCAACCGT GATGACATCA  
 8881 AGATTAACAC CTCAAGGACG TTGAGAAGCC TTGCTTACAA AGGATCTT CTCTCGCGA AAATGCGCTCC AGGGACACCG CGTAAGTGG TAGCAACTCA CGAACTGTC  
 9001 GTACACTGG CGCCGAGATA AAACCAAAAT TTGTGGAGG CGGAAAGATAT GATCTACCG CGTGTGAGG TAAAGGAAAG CGACCGTCA AGAACCTACG GTGACCGAG  
 9121 OCTACATCAC TATOCACAG CGGGGACCCG AGCGTATACG ATGAGATCAT CGGGGAAAGT TTACACAAAG CGGGGACATG CGGAGGACAT TACGTATGAG TGCAAGTGG  
 9241 CGGACTACAC GACCGAACCG GTTCCAGCC CGACCGAAAT CACTGTGAC ACCGCTACAT CGGAGTGGT CGCTTATAAG AGCGAACCAA CGAAGTGGT CGTCAACTCA CGGGACTTGA  
 9361 TCAGACATCA CGACCACAC CGCCAAAGGGA AATTGCAATT CGCTTCAAG TTGATCGGA GTACCTGAT GTGCTTGTG CGCCACCGCG CGAATUTAAAT ACATGCTTAA AACACATCA  
 9481 CGCTCCAAATT AGATACAGAC CACTTGACAT TGCTCACCAC CGGGGAGTA CGGAAACCCG CGGAAACACG CACTGAATGG ATGCGTGGAA AGACCGTCA AGAACCTACG GTGACCGAG  
 9601 ATGCGCTGAA ATACATATGG CGGAAATCATC AGCGGATGG CGCTGATOCG CGAAGTCAAC CACCGGGAGA CGCTTACCGA TGCCGACACG AAATAGTACA CGATTACTACG CATGCCATC  
 9721 CTGTTACACG CACTTGGACG GTGCGATCAG CTACCTGGG GATGATGATT CGCTGAAACG TTGAGTGGT ATGCGCTGTT AAACGGGGCCG GTGAGTGGT CGCCCTGGCGC  
 9841 CGAAACCGCT AGCTCCACG TTGTGCGAC TCTTGTGCTG CGTAAAGTGG CGGAACTACG AAACGTTACG CGGAGGCGT AGTACTGTT GTGCGAACAG TGACCGCTT TCTGGGTGCG  
 9961 AGTGTGTCAT AGTGTGGCG CGTGTGCGC TTCTAATGCG TGCTGCTGC CGTGTGCTG CGTGTGCTG CGAAGTGGAGA CGGCTTACGGAA CATGCCACCA  
 10081 CTGTTACAAAT TTGCGACAG ATACCGTATA AGGCGCTTGG CGGAAAGGCA CGGAACTTGG CGAATGTTCT CGGAGGTTT CGCTTACCC AACGAAAGT  
 10201 ACATTTCTG CAAATTCACCC ACTGTGTCG CGCTCCCAAAT AAACAAATGCG CGCGCTGCTG CGGCGGCGT CGTGTGCGCT CGGCGGCGT CGTGTGCGA CGGGGCTTAC  
 10321 CCTTTATGCG CGGAGGAGCG CAATTTTTCG CGGAGCTGCA GAACGGACCG ATGAGTGGG CGTACGTCGA ACTGTCAGCA CGTGTGCGT CGTACCCACG CGACGGGATT AAAGTGCAC  
 10441 CTGCGCGAT GAAAGTAGGA CTGCGTATAG TGTCGCGGAA CACTACCGT TTCTGAGAT TGTCGCGA CGGAGTCAAC CGGGACACG CTAAAGACTT GAAAGTCTATA GTOGACCA  
 10561 TTTCAGGATC GTTACCGCCA TTGCGATATA AGGTCGTTAT CGATCGCGG CGTGTGCGTAC ACTGTCAGT CGGCGGATAT CGGAGGAGC ATTCGACGTA  
 10681 CCTGTTACAC TGGCAAGGAT CTACCGCCA CGACAGACAT TAGGCTACG AGGCTTCCG CGAAGACCG CGATGTGGG TACACGGG CGCAGTCAAGG ATTGAGTACG TGAAAGAAAC  
 10801 ATCGACCGG CGCACTCGAC GAAACCGCG TTGCGCGTAA TGAGTGGCA CGGAGGCGT CGACTGTTCA CGGCGGAGA CGGCGGCGT CGGCGGCGT CGGCGGCGT  
 10921 CCTTTATGCG GACATCGAG CGACCGCTGG TTCTAAGCGT CAATGTTGG CGTACTGAGT CGACTGAGT CGACTGAGT CGGAGACCTC CGGGGGATGG CGACCGCTGCA GTATGATGCG CGGGGGAG  
 11041 GTCAATGCC CGTACATTCG CATTGAGCA CGGGGAGT CGGAGGCGT CGACGACAT CGTGTGAGGA AGGAGGCGT CGAGTACACT TTACGACCG CGTGTGCGAC CGGAGGCTTGA  
 11161 TCGTGTGCGT GTGCGGGAG AGAGCAACAT CGGAGCAGA ATGAAACCA CGGAGGCGT CGACTGAGT CGGAGGCGT CGGAGGCGT CGGAGGCGT CGGAGGCGT  
 11281 CATGGAGTGG CGTGTGCGT CGGAGGCGT CGACTGAGT CGGAGGCGT CGGAGGCGT CGGAGGCGT CGGAGGCGT CGGAGGCGT CGGAGGCGT CGGAGGCGT  
 11401 ATCGGACCAAG CGAAACCTGCA TTGATCTCG AGGAACTGAT GTGCGATAG CGTACGGCG TGACATTAGA CGGCGGCTTA CGGCGGCAA TATGCAACG CTAAAGACTC GATGACCTTC  
 11521 CGAGGAAGCG CGTGTGCGATAA CGTGTGCGAG CGTGTGCGACA TAACGACAT ATTACACATT TATCTAGGG CGGCGGAAAG CGGCGGAAAG CGGCGGAAAG CGGCGGAAAG CGGCGGAAAG  
 11641 CGCGCTGCG ATAACCTTAA TTATCTTCTT TATTAACAA CGAAATTTTG TTGTTAACAT TTC

Fig. 6B